A Mutation in a Mouse J-protein that Functions in the Diphthamide Biosynthesis Pathway

Thomas Robert Webb

Submitted for the degree of PhD
University of Edinburgh
2005

Acknowledgements

I've been in Edinburgh for just over four years and there are a lot of people who've helped me to get where I am now – writing the last page of my thesis, whahey!!

First I need to thank Ian, who has now been my supervisor for nearly 3 and a half years. Its been a great project and I've really enjoyed my time in the lab, and the nights drinking, we'll have to do the Hallion again sometime. I'd like to thank everyone else in the lab, past and present for all the help they've given me. So far all things science over the past 3 or so years I need to thank Sally for her supervision, and Alan, Pleasie, Katrine, Peter, Margaret, and all the people who've left; Ian, Darren, Sapna, and almost Tom. Special thanks to Lisa for all the mouse stuff and DNA making. There must be other people who I've forgotten, but I've only had an hours sleep in the last couple of days, and that was on the office floor. I know I'll be back up again soon and I'll see you all then (I need to sort everything out, and tidy things up – I'm not sure if I've got enough time to tidy my desk/bench/freezer/other stuff).

Thanks to all the other students and other heavy drinkers for all the drinking nights, especially Dan, for buying me an engagement ring (for Anna, I did pay him back, but there were funny looks in the jewellers). If anyone's down in London please get in touch with me and Anna and we'll show you the sights and smells.

I'd like to thank my parents for so much help over the past few years, both for their personal and financial support, and for their help with our flat. After 8 years of student hardship I'm finally starting a proper job (in 5 days time, for anyone else reading this, some good advice would be to arrange a holiday and not a job for your first week after finishing).

I have to save my biggest thanks for Anna, who I met as soon as I arrived in Edinburgh, and is the best thing that could have happened to me in my time here. I don't think I would have had any chance of finishing this thesis on time without her, or being organised to have a job arranged

(constant nagging "constructive panic" is very effective). We've both been through the same thing in writing our theses, and we've coped remarkably well, however, housework should be done by both of us (especially an aversion to washing up).

I've had a great time in Edinburgh, I'll miss the place and everyone I've met here, (although most have headed for London), so thanks to everyone for making the last 4 years so enjoyable. Next time I'm up I'm sure I'll see you all in t'pub.

I'm sorry if I've forgotten to specifically thank anyone else, but I've been in work for over thirty hours, and I'm getting a bit tired, which probably explains much of the waffle.

Declaration

The experiments described in this thesis were the unaided work of the author except where acknowledgement is made by reference. No part of this work has previously been accepted for any other degree, nor is any part of it being submitted concurrently in candidature for any other degree.

Thomas Robert Webb
September 2005

Abstract

Chromosomal deletions can be used in mutagenesis strategies that screen for recessive phenotypes. Several mouse chromosomal deletions are available which remove the *Pax6* gene and result in the small-eye phenotype. Deletion of the syntenic region in humans results in the phenotypically variable WAGR syndrome. The $Pax6^{Sey\text{-}1H}$ deletion spans 3 Mb and includes 15 genes. This deletion has been used in an ENU mutagenesis screen for phenotypes revealed when hemizygous with the deletion. One such mutation, which is lethal against the $Pax6^{Sey\text{-}1H}$ deletion, has been found from 233 pedigrees. Mice heterozygous for the mutation are phenotypically normal. When homozygous the mutation is generally lethal by E14, and embryos have retarded growth from as early as E9. However, some homozygous mice are born which are small and have extra digits on one or both hind limbs. Sequencing of the genes in the deletion interval identified a splice site mutation in *Dnajc20* a member of the J-domain family of molecular chaperones. The mutation causes the in frame skipping of exon 4, leading to an mRNA that encodes a protein with an internal 23 amino acid deletion.

The yeast orthologue of *Dnajc20* has recently been identified as one of five genes required for the biosynthesis of diphthamide, a post-translational modification that is unique to translation elongation factor 2 (eEF2), the target of diphtheria toxin and *Pseudomonas* exotoxin A. Here it is shown that the mutation in the mouse Dnajc20 protein results in the abolition of diphthamide on eEF2 in these animals. Additionally mice mutated for *Dnajc20* show remarkable similarity to the phenotype of *Ovca1* mutant mice. *Ovca1* is a tumour suppressor whose yeast orthologue was also identified as being required for diphthamide biosynthesis. The phenotypes of these two mutations suggest that the diphthamide biosynthesis pathway, which is conserved from archaebacteria to eukaryotes but absent in eubacteria, is essential for normal mammalian development and survival. Furthermore the human orthologue of *Dnajc20* is suggested as the gene responsible for the manifestation of polydactyly in some WAGR patients.

Contents

Chapter 1

Introduction

Chapter 2

Region based ENU mutagenesis identifies a mutation in a J-protein

Chapter 3

Dnajc20 and the mouse family of J-proteins

Chapter 3

Dnajc20 and the mouse family of J-proteins (Continued)

Chapter 4

Dnajc20 is required for diphthamide biosynthesis

Chapter 5

Analysis of a potential homologous mutation in humans

Chapter 6
Discussion

Chapter 7

Materials and Methods

Chapter 7

Materials and Methods (Continued)

Figures and Tables

List of Abbreviations

°C	degrees centigrade
aa	amino acid
AdoMet	S-adenosylmethionine
bp	base pair
BSA	bovine serum albumin
cDNA	complimentary deoxyribonucleic acid
DNA	deoxyribonucleic acid
dNTP	dinucleotide triphosphate
ds	double stranded
DT	diphtheria toxin
E	embryonic day
EMS	ethylmethane-sulphonate
ENU	*N*-ethyl-*N*-nitrosurea
ES	embryonic stem
EST	expressed sequence tag
ETA	*Pseudomonas* exotoxin A
g	gram
kb	kilobase pairs
kDa	kiloDalton
kti	killer toxin insensitive
l	litres
lacZ	*B-galactosidase* gene
LOH	loss of heterozygosity
M	molar
m	prefix *milli*
μ	prefix *micro*
mRNA	messenger RNA
n	prefix *nano*
nt	nucleotides
OD	optical density
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
rRNA	ribosomal RNA
RNA	ribonucleic acid
RNAi	interfering RNA
RT-PCR	reverse transcription-polymerase chain reaction
TBE	tris, boric acid, EDTA
TE	tris, EDTA
tRNA	transfer-RNA
WAGR	Wilms' tumour, aniridia, genitourinary abnormalities, mental retardation

Chapter 1

Introduction

1.1 Introduction

The past ten years has seen the sequencing of the genomes of humans[1,2] and an increasing number of model organisms[3-6]. This has been a landmark achievement in biology, providing a mass of information concerning genome structure and content. The challenge now is functional genomics, the determination of a biological role of all the genes in the human genome. Evidence of gene function can be obtained by studying the expression of genes and their distribution in normal tissues. However, the most powerful tool in the dissection of gene function is through generation of mutations and their subsequent characterisation. The use of mutation in the identification of gene function will be achieved through the use of model organisms. The major model organisms including, yeast (*Saccharomyces cerevisiae*), fruitfly (*Drosophila melanogaster*) (referred to here as *Drosophila*), worm (*Caenorhabditis elegans*), and the mouse (*Mus musculus*), are proven tools for the study of gene function and the publication of the genomes for each of these organisms has increased their effectiveness in functional genomics. Each of these models already has a substantial mutant resource and programmes are being undertaken which both increase and advance the information available.

1.2 Eukaryotic mutation and conservation

The mouse has emerged as the best model for the study of human gene function and study of the pathogenesis of human disease. This is largely due to the fact that of the advanced model organisms, it is the mouse that is most similar to humans. However, the mouse has a relatively small mutant resource, comparing the number of mutations to the number of genes. The study of mutations in the mouse has been performed for as long as it has been a model organism. It was mouse fanciers, breeding coat colour variants at the turn of the last century, who were essentially responsible for the start of mouse genetics and many spontaneous mutations in the mouse have now been maintained for many years. The use of radiation to mutagenise the mouse in the mid-20th century rapidly increased the mutant resource. However, it is only relatively recently, with the large increase in markers and the genome project, that large-scale mutagenesis of the mouse has become viable, and the aim of creating at least one mutation per gene a possibility. The use of mutagenesis in the mouse will be introduced more fully in Chapter 2.

It is almost 10 years since the yeast genome was published, bringing with it the possibility for vast, genome wide, gene function studies[3]. The aim of creating one mutation per gene has been achieved in yeast, where close to a comprehensive collection of yeast deletion strains has been produced[7]. The yeast has already been a very successful model in the dissection of fundamental biological processes, for example

many of the genes regulating cell cycle were identified as cell division mutations in the different yeast species[8-10]. Now, large-scale phenotype screens make use of the global deletion resource, identifying changes in growth rate due to the presence of stressors or chemicals, in the dissection of particular processes, or even in a search for human disease genes[11-13]. A genome-wide mutation resource has other uses besides global screens. Small mutagenesis screens are still performed with the deletion strains providing a quick resource for confirming the mutated loci. The deletion strains are also used for the study of mammalian genes. Human or mouse genes can be used to replace a yeast orthologue in investigations of functional equivalence, a powerful tool in the identification of gene function.

In *Drosophila*, phenotype-driven mutagenesis has been enormously successful in the identification of genes required for embryonic development[14]. Large-scale mutagenesis screens have been performed using either the chemical mutagen EMS (ethylmethane-sulphonate) or P-element transposons. The screens have been particularly efficient at identifying genes in a particular pathway, for example, the components of the Wnt signalling pathway were first identified by mutagenesis of *Drosophila*[15]. Modifier screens, where mutations are sought that alter an already sensitised background, have also been efficient in *Drosophila*, specifically with the identification of genes relevant to human neurodegenerative disorders[16]. A problem with the phenotype-driven screens in *Drosophila* has been the mapping of EMS-induced mutations. Only recently, with the

identification of markers, and the ability to map mutations to the genome sequence has the usefulness of phenotype driven screens in *Drosophila* been improved.

The large number of mutations generated by genotype-driven screens, such as the yeast deletion resource[7], and the RNAi approach in worm, which covers the majority of the worm genes[17], as well as the easier identification of mutations induced in phenotype driven screens means that there is now a vast eukaryotic mutant resource.

The genome sequencing effort of recent years has allowed a vast increase in the identification of conserved gene families both within a single species and across multiple species. This ability to compare evolutionary conserved gene families among the different model organisms is one of the major benefits of the post-genomic era[18]. The Gene Ontology consortium has assigned standard annotations to gene products, which describe the biological process, the molecular function and the cellular component or location, based upon the characterised roles of genes within conserved families[19].

The public availability of information allows a large amount of information about gene function to be quickly identified. The presence of large amounts of conservation between the genes of different eukaryotes means that the investigation of a particular gene can be investigated throughout multiple species.

1.3 Project outline

A combination of mutation and conservation provides the basis for the work detailed within this thesis, from the isolation of a developmentally relevant gene by mutagenesis of the mouse, to the identification of its biological role by inference from its recently identified yeast orthologue.

1.3.1 Mutagenesis in the mouse

Our group has recently performed a small ENU mutagenesis screen targeted to a region of mouse chromosome 2 around the *Pax6* and *Wt1* genes. Chapter 2 introduces the use of mutagenesis in the mouse and gives the background to this screen. The screen isolated a mutation that results, predominantly, in embryonic lethality. The embryos are small, developmentally delayed and have polydactyly of the hind limbs. The mutation, by the nature of the screen, was localised to a particular chromosomal region. Sequencing of the genes within the region identified a mutation in a functionally novel J-protein. The orthologous human protein, along with the rest of the chromosomal region is syntenic with human chromosome 11p13. Deletion of this region in humans results in WAGR syndrome (Wilm's tumour, Aniridia, Genitourinary abnormalities and mental Retardation). Recently WAGR patients have been identified with very similar polydactyly to that observed in the mutant mice[20,21]. Chapter 5 describes the human

syndrome in more detail, as well as an investigation as to whether the orthologous human J-protein is the cause of the polydactyly in these patients.

1.3.2 J-proteins; homologues and orthologues

The mutated J-protein has previously been named *mmDjc7* in mouse[22], however, the mammalian J-protein family has been complicated in recent years by a number of changes of nomenclature. Chapter 3 describes the identification of mouse J-proteins and the clarification of the mouse J-protein family. We therefore suggest the name *Dnajc20* for this particular gene, in line with the current nomenclature. The bacterial DnaJ protein provides the name for the J-protein family, a large family of molecular chaperone proteins that are conserved throughout prokaryotes and eukaryotes. Chapter 3 provides a detailed introduction to the J-proteins, as well as their chaperone partners, the Hsp70s.

The Dnajc20 protein is conserved throughout eukaryotic evolution (See Chapter 3), and it was the identification of these sequence orthologues that consequently resulted in the identification of its biological role.

1.3.3 Diphthamide biosynthesis

Recently, through the investigation of mutations in yeast that result in resistance to diphtheria toxin, Liu *et al* identified the yeast orthologue of Dnajc20 as one of the proteins required for the synthesis of diphthamide[23]. Diphthamide is a post-translational modification that is only found on translation elongation factor 2 (eEF2)[24]. ADP-ribosylation of diphthamide by diphtheria toxin or *Pseudomonas* exotoxin A results in the inhibition of protein synthesis and cell death. The diphthamide residue was originally identified 25 years ago[25], however its function in normal cellular physiology is still unknown. Chapter 4 introduces this process and describes the investigation of diphthamide in the mutant *Dnajc20* mice. Another of the factors required for diphthamide biosynthesis in yeast is the orthologue of mouse *Ovca1*[23]. This gene has recently been knocked out in the mouse[26] and has a very similar homozygous phenotype to the *Dnajc20* mutant mice. Together these two mice show a requirement for diphthamide in normal mammalian development and postnatal survival. Furthermore, the *Ovca1* heterozygous mice spontaneously develop tumours, potentially linking diphthamide and *Dnajc20* to tumourigenesis.

Chapter 2

Region based ENU mutagenesis identifies a mutation in a J-protein

2.1 Introduction

The availability of the complete sequence of both the human[1,2] and mouse genomes[6] has revealed the volume of the coding potential of the mammalian genome. The task ahead is to decipher the diverse function of the estimated 25-30,000 genes. In an effort to meet this challenge there has been an escalation in the application of mutagenesis screens in the mouse. These studies aim to provide an insight into the function of a gene via the result of its mutant phenotype.

2.1.1 Mouse mutagenesis

There are essentially two mutagenesis schemes that can be employed in the production of mouse mutants; genotype-driven and phenotype-driven. Genotype-driven mutagenesis of the mouse involves the construction of transgenic lines, whereby the gene of interest is knocked out or is disrupted by a knock in. This strategy requires an interest in a particular gene and is generally performed on those genes that have been very well studied, meaning that novel loci are ignored. Gene-trap mutagenesis provides a genotype-driven strategy for the systematic genome wide screening for mouse mutations. This mutagenic approach involves the insertion of marker transgenes into the genome of ES cells. If this marker integrates within an active gene, that gene becomes disrupted and is tagged by the marker, allowing straightforward identification of the mutated loci. Theoretically, the random disruption of genes in a gene-trap mutagenesis

screen could work as a phenotype driven approach. However, the easy identification of the mutated gene and the cost of going from an ES cell to a mouse mean that a mutation remains in ES cell form until it becomes interesting to an investigator, and therefore, is used in a genotype-driven manner.

Phenotype-driven mutagenesis of the mouse involves the use of the chemical mutagen *N*-ethyl-*N*-nitrosourea (ENU), which randomly generates single base pair changes throughout the genome. This approach makes no prior assumption concerning the nature of the genes involved with the interest being confined to a particular phenotype. A large number of mice are screened for the phenotype of interest, and following the identification of such animals, the mutated loci can be mapped by breeding the animals and utilising the genetic markers available. It is the time and cost of this mapping that is the major disadvantage of phenotype-driven compared to the genotype-driven screens.

The major difference between a gene-trap based genome wide screen and an ENU mutagenesis strategy is the type of mutation generated. A gene-trap disrupts a gene resulting, in the production of a null allele, in the majority of cases. ENU mutagenesis causes a greater variety of mutations ranging from null alleles to hypermorphs. This range of alleles provides the chief benefit of ENU mutagenesis, as it gives a better reflection of the mutations causing human disease, and are therefore the best way to dissect the pathogenesis of human conditions. Furthermore, point mutations allow a series of mutations to be generated within the same gene, which may give different

phenotype and allow the identification of functionally relevant features within a gene product.

2.1.2 Mode of action of ENU

ENU is an alkylating agent that predominantly induces point mutations via the transfer of its ethyl group to oxygen and nitrogen radicals at a number of sites within the nucleotides; the O2, O4, or N3 of thymine, N1, N3, or N7 of adenine, O2 or N3 of cysteine, and O6, N3, or N7 of guanine[27,28]. The presence of the ethyl group on a nucleotide is called a DNA adduct, and this structure can result in mispairing during the next round of DNA replication, leading to a base pair substitution. The most common ENU generated mutations in the mouse are AT to TA transversions, and AT to GC transitions[29], and are the result of mispairing adducts at O2 and O4 of thymine[28].

The mutagenic action of ENU is most potent in male spermatogonial stem cells[29]. ENU is administered by intraperitoneal injection, which due to its cytotoxicity, renders the mice sterile. Fertility recovers by repopulation from stem cells that harbour mutations.

Specific locus experiments indicate the mutation rate in mouse using ENU is in the region of 10^{-3} per locus per gamete, meaning that for every 1000 gametes from a mutagenised male there may be a mutation in any gene of interest[30]. However, the rate of mutation in any particular gene can vary significantly, depending on various factors,

including; the size of the gene, the number of splice sites and the coding length, the presence of domains or regions important for the overall structure and function of the protein, and potentially the presence of hot spots in certain regions of the genome. The mutation rate at the genome level has been shown to be around one base change per megabase of DNA[31].

2.1.3 ENU Mutagenesis Strategies

The efficiency of ENU, as well as its low cytotoxicity, makes it the most useful mutagen for phenotype-driven screens in the mouse. A number of approaches have been used in the generation of mutant animals, but the method used essentially depends on whether it is a screen for dominant or recessive mutations. Dominant screens have a simple genetic design, whereby male mice are injected with ENU and then mated with untreated females. The progeny from this cross are the G1 test class, and these are screened for the phenotype of interest. Affected animals are bred to confirm the heritability of the abnormality, and intercrossing allows the homozygous phenotype to be studied. Figures for dominant screens suggest a heritable mutant phenotype is present in at least 2% of the G1 animals[32,33].

Recessive screens require a more complex strategy for the recovery of the test class. The first step is the same as in the dominant screens; mutagenised males are mated with untreated female mice to produce the G1 progeny. Each G1 mouse represents an

individual mutant line that may be heterozygous for many recessive mutations. These G1 animals are mated with wild-type mice to produce a G2 progeny, with a 50% chance of inheriting any recessive mutations from the G1 founder. Backcrossing these G2 animals to the G1 generation will result in G3 progeny that can undergo phenotypic analysis. Of the G3 progeny 25% will be homozygous for the particular mutation. Figures for recessive screening strategies suggest that up to 30% of the G3 animals will carry a heritable mutation[34].

2.1.4 Region based screens

The production of mutations using genome-wide recessive screens is slow and time consuming. However, recessive mutations can be recovered more quickly in the particular regions of the genome that are covered by deletions and inversions. In these screens the G1 progeny from ENU treated males are mated to mice that carry a deletion that covers a defined chromosomal region. This will result in a phenotype due to hemizygosity in the mice that inherit the ENU induced mutation from one parent and the deletion from the other parent. A benefit of this strategy is the immediate localisation of a mutation to the chromosomal region covered by the deletion, avoiding the need to map the mutation.

Regionally directed mutagenesis screens have been performed against well-defined genetic intervals and generally employ visible marker loci. For example, a mutagenesis

project directed against the albino, *Tyr*, locus on mouse chromosome 7 has been undertaken[30]. This screen isolated 13 mutations from 972 gametes with phenotypes ranging from embryonic lethality and runting, to neurological defects. The strategy that Rinchik *et al* employed in their screen involved the mutagenesis of male mice homozygous for the recessive albino mutation in tyrosinase. These mice were mated with wild-type females to produce an F1 progeny that were heterozygous for the albino marker and had a heterozygous mutagenised genome, with the possibility of a mutation in the vicinity of the albino locus. The G1 progeny were then mated with a deletion-carrying class that were hemizygous for the chinchilla mutation in tyrosinase. The use of these coat colour mutations allowed the quick scoring of the G2 progeny; with the test class (those with a possible ENU induced mutation against the deletion) were albino, and the carriers of any induced mutation were light chinchilla (due to the compound albino/chinchilla genotype). The other G2 progeny were wild-type.

ENU mutagenesis screens such as that performed by Rinchik *et al* have many benefits; they allow the generation of recessive mutations over two generations, the scoring of potential mutations by phenotyping, and any mutation will be mapped to a relatively small region[30].

2.2 Background

2.2.1 A region based screen in the mouse *Pax6* region

Our group, in collaboration with MRC Harwell, has recently completed a small ENU screen for recessive mutations directed against the region of mouse chromosome 2 around the *Pax6* gene. The region is of interest for several reasons; firstly there are a number of mouse mutations, the *Small eye* (*Sey*) alleles, which have been well characterised. These mutations include point mutations within the *Pax6* gene, as well as deletions that remove up to several megabases of DNA. These deletions, which have made the screen possible, are associated with a number of growth defects when heterozygous, and cause early lethality when homozygous. Secondly, the syntenic chromosome region in humans, 11p13, is deleted in the phenotypically variable WAGR syndrome, consisting of Wilms' tumour, aniridia, genitourinary abnormalities, and mental retardation (this human syndrome will be discussed in more detail in chapter 5). Finally, the phenotypes of the deletions in the mouse and humans are not fully explained by the individually mutated genes, suggesting the presence of further loci within this region that are important for growth and survival, as well as the cause of the early embryonic lethality.

2.2.2 The *Pax6* region

The *Pax6* region on mouse chromosome 2, along with the syntenic region on human chromosome 11p13, is a locus defined by mutation; the *Small eye* alleles in mouse and WAGR syndrome in humans. The *Small eye* phenotype resembles human aniridia, and displays a similar pattern of inheritance, thus providing a mouse model for a human disease[35,36]. The mapping of WAGR chromosomal markers to the region of mouse chromosome 2 near the *Small eye* locus, revealed that not only did the *Sey* alleles provide a model for aniridia; the deletion alleles could also represent a model for WAGR syndrome[37]. The subsequent identification that human *WT1* and *PAX6* genes were responsible for Wilms' tumour and aniridia respectively led to the identification of the genes in the mouse[38,39].

The first of the *Small eye* alleles arose spontaneously and is due to a point mutation in *Pax6*, resulting in premature termination before the homeobox domain and loss of function of *Pax6*[38,40]. The $Pax6^{Sey}$ mutation causes small eyes, cataracts, and no iris in heterozygous animals. Homozygotes are born with no eyes or nose and die soon after birth[35,36,40]. Several other point mutations have been generated in *Pax6* by ENU mutagenesis. The Neuherberg series of *Pax6* alleles include 8 nonsense mutations, which would result in truncated proteins, all of which have phenotypes suggesting null alleles[41,42]. Of the other $Pax6^{Neu}$ alleles, $Pax6^{4Neu}$ has a missense mutation in its homeodomain, and $Pax6^{7Neu}$ is a result of a change in the Kozak consensus sequence.

The first deletion alleles to be identified were *Pax6*$^{Sey\text{-}Dey}$ and *Pax6*$^{Sey\text{-}1H}$, both of which remove the *Wt1* and *Pax6* genes[43,44]. The heterozygous and homozygous phenotypes of mice carrying these deletion alleles are more severe than the *Pax6*^{Sey} point mutation. The *Pax6*$^{Sey\text{-}Dey}$ allele arose spontaneously and the *Pax6*$^{Sey\text{-}1H}$ allele was induced by the X-ray irradiation of oocytes. Animals heterozygous for these two alleles are 10-20% smaller than their wild-type littermates and have reduced viability. The most obvious characteristic of these animals is a white spot on their belly. There is prenatal loss of around two thirds of deletion heterozygotes, and homozygotes die around preimplantation[43,45]. Compound *Pax6*^{Sey}/*Pax6*$^{Sey\text{-}1H}$ heterozygotes display a more severe phenotype than the *Pax6*^{Sey} homozygotes, being smaller throughout embryonic development and more retarded in facial development.

The *Sey* deletion alleles do not replicate the Wilms' tumour or genitourinary abnormalities caused by the heterozygous loss of the *WT1* gene in humans, even though in both the *Pax6*$^{Sey\text{-}Dey}$ and *Pax6*$^{Sey\text{-}1H}$ deletion carriers *Wt1* is deleted. These findings suggest that mouse development is not as susceptible to *Wt1* loss as human development[37]. Targeted mutagenesis has been used to create null *Wt1* mice[46]. The *Wt1* heterozygotes are normal, but the homozygotes die at around embryonic day 15, revealing that in mouse the loss of one copy of *Wt1* is not sufficient to replicate the human abnormalities.

The increased severity of the *Sey* deletions compared to loss of *Wt1* or *Pax6* in mouse, and as the additional features of WAGR syndrome (particularly mental retardation), has led to a lasting interest in the region. However, only a small number of other genes have been identified between *Wt1* and *Kcna4*, the genes representing the end points of the $Pax6^{Sey\text{-}1H}$ deletion[47]. Reticulocalbin, a Ca^{2+} binding protein of the endoplasmic reticulum, has been localised to the region between *Wt1* and *Pax6*. Another gene, *Elp4* (originally *Paxneb*), was identified immediately distal of *Pax6*[48]. Interestingly, intron 9 of this gene was shown to contain control elements of *Pax6*, which also result in aniridia when mutated. The only other known genes that have been localised to this region were *Fshb* and *Kcna4*, which are towards the distal breakpoint of the $Pax6^{Sey\text{-}1H}$ deletion. However, these genes are not removed by the $Pax6^{Sey\text{-}Dey}$ deletion. The similar phenotypes produced by $Pax6^{Sey\text{-}Dey}$ and $Pax6^{Sey\text{-}1H}$ deletions suggest that loss of either *Fshb* or *Kcna4* does not result in any overt phenotype[47]. Therefore, there must either be novel mutations in those genes already identified or other novel loci responsible for features of the *Sey* deletion alleles and WAGR syndrome.

2.2.3 The strategy of the screen

The *Sey* deletions on mouse chromosome 2 are appropriate tools for performing a region based ENU mutagenesis screen. The recessive screen provides a platform for the identification of developmentally relevant genes, as well as the further dissection of the deletion phenotypes that currently cannot be explained totally by the known genes in the

region. The synteny between this region of mouse chromosome 2 and the human WAGR region implies that any mutations identified will have relevance to the human syndrome.

Similar to other region specific screens, visible marker loci are utilised in the form of the *Small eye* mutations to track the mutagenised genome over two generations. However, unlike other region based screens there are no viable recessive alleles available to use as a marker (such as the albino mutation used by Rinchik *et al*[30]). Instead this breeding strategy involves the use of a heterozygous marker mutation, the $Pax6^{Sey\text{-}Neu}$ allele. This mutation is semi-dominantly inherited, with heterozygous mice displaying micropthalmia. The mutation is embryonic lethal when homozygous, therefore, no novel alleles of *Pax6* will be recovered in this screen due to the likely embryonic lethality of any compound heterozygous progeny of the first cross.

The screen was originally designed against three of the *Small eye* deletion alleles, $Pax6^{Sey\text{-}1H}$, $Pax6^{Sey\text{-}2H}$, and $Pax6^{Sey\text{-}3H}$. The use of multiple deletions allows the further mapping of mutations via complementation testing. However, the low fitness and fecundity of the larger of these deletions, $Pax6^{Sey\text{-}2H}$ and $Pax6^{Sey\text{-}3H}$, meant that they were uneconomical for use in screening. Therefore, only the $Pax6^{Sey\text{-}1H}$ allele was used.

The scheme for the region-based screen for recessive mutations is shown in figure 2.1. ENU treated BALB/c males are mated with females heterozygous for the $Pax6^{Sey\text{-}Neu}$

mutation; producing G1 progeny with a heterozygous mutagenised genome against the *Pax6* marker allele. G1 $Pax6^{Sey\text{-}Neu}$ heterozygotes are selected, as these identify the animals that potentially carry an ENU-induced mutation in the region of interest inherited from the mutagenised BALB/c male. The $Pax6^{Sey\text{-}Neu}$ heterozygotes are crossed with a deletion carrier to produce four theoretical G2 progeny classes; a potential carrier of an ENU induced mutation, the potential mutation against the deletion (the test class), a $Pax6^{Sey\text{-}Neu}$ carrier, and the theoretical $Pax6^{Sey\text{-}Neu}/Pax6^{Sey\text{-}1H}$ compound mice. This $Pax6^{Sey\text{-}Neu}/Pax6^{Sey\text{-}1H}$ compound genotype is lethal and therefore, will not be seen in the F2 progeny.

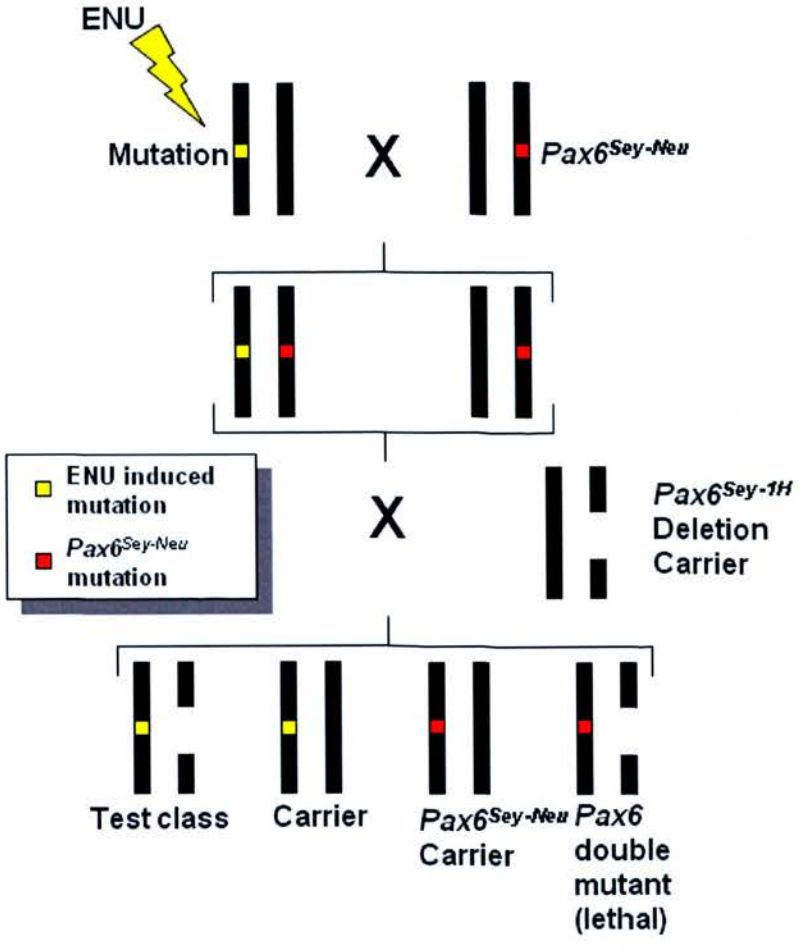


Figure 2.1 ENU mutagenesis screen against the *Pax6* region

Male BALB/c mice were treated with ENU, possibly inducing a mutation in the desired region (yellow box), and mated with heterozygous $Pax6^{Sey\text{-}Neu}$ (red box) females. The progeny of this cross that carried the $Pax6^{Sey\text{-}Neu}$ mutation were investigated for the presence of a heritable mutation on the other chromosome by crossing to the $Pax6^{Sey\text{-}1H}$ mice. The progeny of the this cross contains the test class, which is identifiable by the $Pax6^{Sey\text{-}1H}$ mutation. If no $Pax6^{Sey\text{-}1H}$ mice are born, a recessive lethal mutation has been induced.

The *Pax6*$^{Sey\text{-}1H}$ carriers are the test class, and can be distinguished from the *Pax6*$^{Sey\text{-}Neu}$ mice by phenotype (*Pax6*$^{Sey\text{-}1H}$ carriers are smaller and have the white belly spot). The *Pax6*$^{Sey\text{-}1H}$ mice can be examined for a modified phenotype, implying the presence of a recessive mutation in the region. If the *Pax6*$^{Sey\text{-}1H}$ class is absent, a recessive lethal mutation may have been induced. Recessive lethal mutations can be retrieved from the carriers. The transmission rate of the *Pax6*$^{Sey\text{-}1H}$ deletion chromosome is 34%, giving expected test class progeny numbers at a ratio of approximately 2x carrier to 1x deletion carrier to 2x *Pax6*$^{Sey\text{-}Neu}$ carrier. This gives a probability of finding no deletion carriers by chance at less than 1% when there are more than 20 offspring.

2.2.4 The MUTS1/14 mutation

One mutation has been retrieved from 233 fully tested lines. This mutation, MUTS1/14, was found to be lethal against the *Pax6*$^{Sey\text{-}1H}$ deletion. Out of 35 F2 progeny there was one animal detected carrying the *Pax6*$^{Sey\text{-}1H}$ deletion chromosome, compared to the expected number of seven. The surviving *Pax6*$^{Sey\text{-}1H}$ deletion mouse might be the result of a "leaky" lethal phenotype due to a penetrance effect. Or, more likely, a recombination event within the deletion region, whereby a cross-over between the *Pax6*$^{Sey\text{-}Neu}$ allele and the induced mutant locus results in the mutation being inherited by the *Pax6*$^{Sey\text{-}Neu}$ carriers instead of the *Pax6*$^{Sey\text{-}1H}$ test class.

Carrier mice from the MUTS1/14 line were intercrossed to characterise the homozygous phenotype. MUTS1/14 homozygotes were occasionally viable, however in the majority of homozygotes the mutation resulted in embryonic lethality (Table 2.1). One fifth of the expected number of MUTS1/14 were born, and of these only three reached weaning, two were dead at birth, and four were euthanised due to poor health by postnatal day 2. The homozygotes that were born were small (Figure 2.2.A), with some having one or more extra digits on either one or both hind limbs (Figure 2.2.B and C). Figure 2.2.D illustrates the preaxial polydactyly of the MUTS1/14 mice. It shows a bone stain of the right hind limb of one of the MUTS1/14 animals that was dead at birth (this experiment was performed subsequent to the work at MRC Harwell, see Materials and Methods). There is total duplication of digit 1 and partial duplication of digit 2. Animals with one extra digit showed duplication of digit one only. This phenotype differs from the more common form of preaxial polydactyly, which involves the "mirror image" duplication of any of digits 2-5.

Table 2.1 MUTS1/14 intercross 1

Genotype	Number	Phenotype
Wild-type	45	Wild-type
Heterozygote	89	Wild-type
Homozygote	9	Small, extra digits on one or both hindlimbs

Table 2.1. Intercrossing the MUTS1/14 carriers. A fifth of the expected number of homozygotes were born. Of these only 3 reached weaning, 2 were dead at birth, and 4 were euthanised.

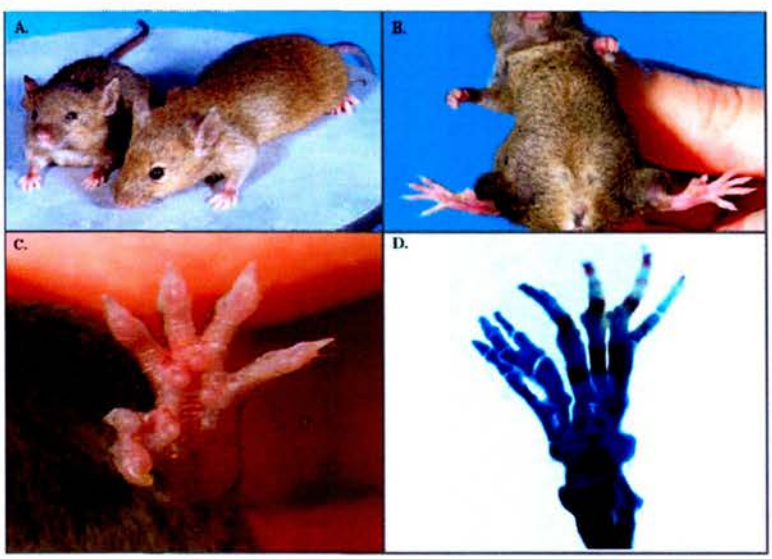


Figure 2.2 MUTS1/14 homozygous mice.

A small number of MUTS1/14 homozygous mice were born in intercrosses performed at MRC Harwell. The MUTS1/14 homozygous mice are smaller than their wild-type littermates. 2.2.A Homozygous MUTS1/14 embryo (left) with a wild-type littermate. 2.2B and 2.2.C, preaxial polydactyly on hind limbs of MUTS1/14 homozygous mice. 2.2.D Alizarin red/alcian blue bone staining of the right hind limb of a MUTS1/14 pup born at MRC Harwell and euthanised at P, showing preaxial polydactyly. There is complete duplication of digit 1, and bifurcation of a second digit.

The MUTS1/14 mutation is known to be in a gene within the *Pax6*$^{Sey\text{-}1H}$ deletion interval, which is approximately 3Mb in size, and includes the genes between *Wt1* and *Kcna4*[47]. However, recombination events during the breeding of MUTS1/14 carriers reduced the size of the candidate interval. Recombination events in two mice resulted in the homozygosity of the mutagenised BALB/c chromosome telomeric of the *D2Mit100* marker. As both mice were phenotypically wild-type, the mutation was localised between the proximal breakpoint of the *Pax6*$^{Sey\text{-}1H}$ deletion and the *D2Mit100* marker. Therefore the MUTS1/14 mutation was confined to a 1.4 Mb region between the *Wt1* gene and the *D2Mit100* marker.

2.2.5 Identification of candidates for the MUTS1/14 mutation

At the outset of this investigation the annotation of this region of mouse chromosome 2 was poor, with only 4 known genes being present in the 1.4Mb region around *Pax6*; the *Pax6* gene itself, *Wt1*, *Rcn* and *Elp4* (originally published as *Paxneb*). At the same time as the ENU screen was undertaken, the UK Mouse Sequencing Project began the high quality sequencing of this region, as well as other sections of chromosomes 4, 13, and X. Using this sequence information I was able to characterise the coding potential of the region between *Wt1* and *D2Mit100* via comparative sequence analysis and gene prediction[49]. This analysis identified the presence of at least 8 genes and 61 exons. Subsequently, a 9 exon gene, *DCDC1*, was identified in humans[50] and is also present in mice. My original analysis had combined the C-terminus of this gene with the N-

terminus of the *4732421G10Rik* gene. *Dcdc1* and *4732421G10Rik* produce related gene products. Both proteins contain one doublecortin domain (DCX), and as the other proteins in this family contain two of these domains, my original investigation resulted in the misannotation of these two genes. The updated list of genes is included in Table 2.2 and Figure 2.3.

Table 2.2 Candidate genes

Gene (Mouse)	Exons	Comments
Wt1	10	Tumour suppressor gene
0610012H03Rik	4	Expressed in the kidney, contains 4HBT domain
Rcn1	6	Ca^{2+}-binding protein of the endoplasmic reticulum
Pax6	11	Paired domain and a homeodomain containing developmental regulator
Elp4	10	Homologue of the yeast elongator complex subunit ELP4
1500034J20Rik	8	Homologue of yeast IMP1
Dnajc20 (*1700030A21Rik*)	5	Homologue of the bacterial heat-shock protein DnaJ
Dcdc1 homologue	9	Expressed in testis and brain, contains doublecortin domain
4732421G10Rik	5	Contains doublecortin domain

Table 2.2. The 9 genes encoded between the proximal breakpoint of the *Pax6*$^{Sey\text{-}1H}$ deletion and the *D2Mit100* marker.

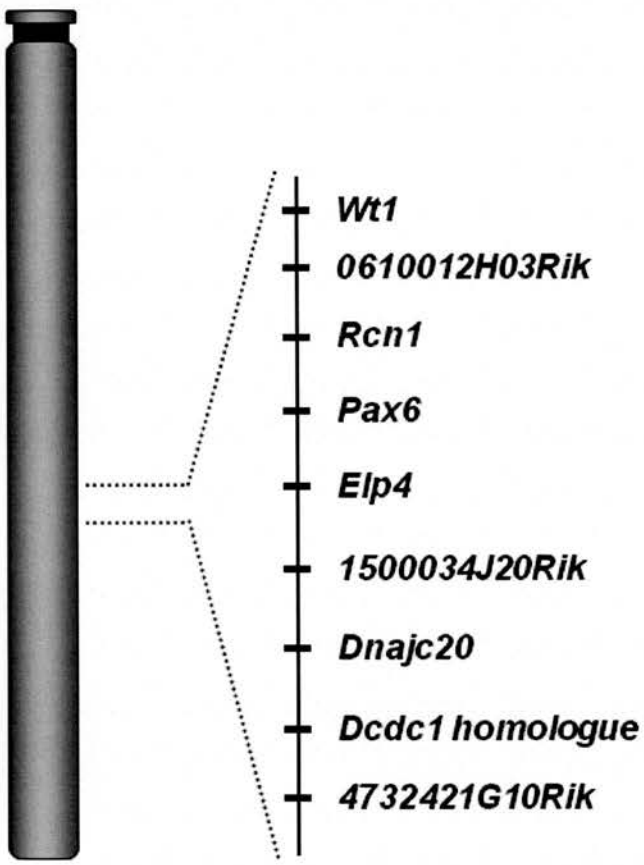


Figure 2.3 Candidate genes for MUTS1/14

The position of the 9 candidate genes for the MUTS1/14 mutation on mouse chromosome 2.

2.3 Results

2.3.1 Identification of a splice mutation in a J-protein

To identify the base change responsible for the MUTS1/14 phenotype I sequenced the exons of several of the genes in the candidate region. As the MUTS1/14 homozygous phenotype is universal growth retardation, the gene responsible would be expected have a wide expression pattern. The first genes sequenced were chosen due to their EST coverage and were *Elp4*, *1500034J20Rik*, *Dnajc20*. Although the *Wt1* null phenotype does not resemble the MUTS1/14 homozygotes, it was also sequenced due to the expression of a *Wt1* YAC transgene between the developing digits of the hindlimbs[51] (however, as expected no mutation was identified in *Wt1*).

The sequencing of the exons of the candidate genes revealed the presence of one base change, a T to A transversion at position +2 of the splice donor site in intron 4 of *Dnajc20* (Figure 2.4.A). The +2 position of a splice donor site is 100 % conserved and is required for correct splicing of a transcript [52]. Mutations at this position in numerous other genes have previously been reported and always result in the skipping of the adjacent exon[52]. To test the effect of the identified base change on the splicing of the *Dnajc20* mutant transcript RT-PCR analysis was performed. Figure 2.4.B shows the amplification of the *Dnajc20* transcript from wild-type, heterozygote, and homozygote RNA between exons 3 and 5.

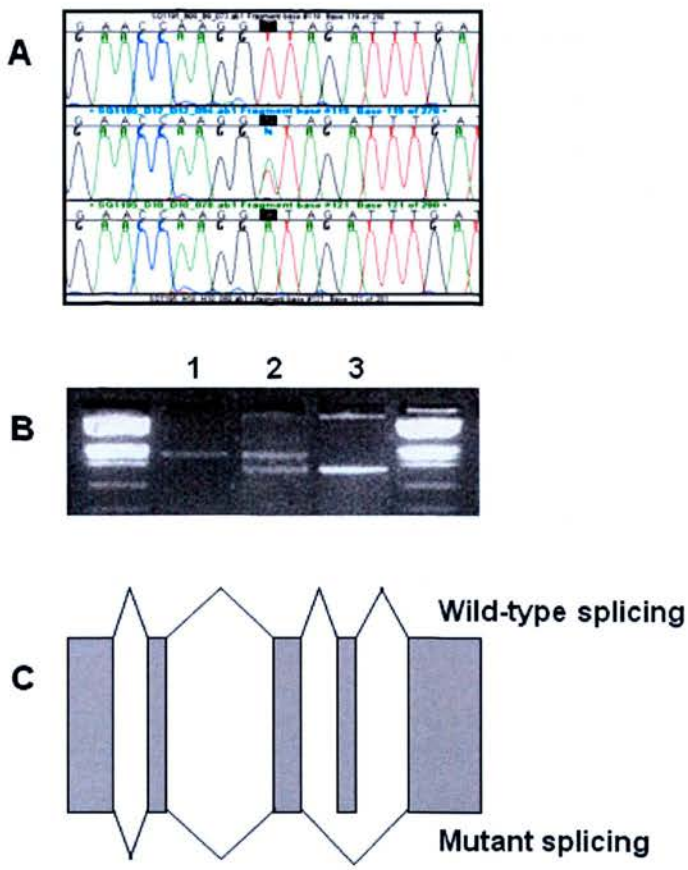


Figure 2.4 A splice mutation in *Dnajc20*

A. Sequence view of the exon/intron boundary, with position +2 highlighted, illustrating the T to A transversion. Top sequence is from a wild-type. wild-type, middle sequence is from a heterozygote, bottom sequence is from a mutant.

B. RT-PCR between exon3 and exon 5 of *Dnajc20*. Lane 1 is wild-type, 2 is from a heterozygote, 3 is from a mutant. The higher band represents the presence of exon 4, the lower band represents the skipping of exon 4. The heterozygote sample contains both bands at approximately equal levels.

C. Schematic of *Dnajc20* splicing in the wild-type and in the heterozygote.

The expected wild-type amplified product from RT-PCR of the *Dnajc20* transcript between exon 3 and 5 is 246bp. This corresponds to the size of the band from the wild-type sample, and the larger band in the heterozygote (Figure 2.4.B). The amplified product of the mutant sample, and the smaller band in the heterozygote sample, is approximately 69bp smaller (Figure 2.4.B), equivalent to the skipping of exon 4. Sequencing of the RT-PCR product of the mutant sample confirmed the absence of *Dnajc20* exon 4 in the MUTS1/14 mutant line. Figure 2.4.C illustrates the splicing of the *Dnajc20* transcript in the MUTS1/14 mutation.

Dnajc20 is a 5 exon gene, with the start site being present in exon 2 and a TGA termination codon in exon 5 (Figure 2.5). The mutation in the MUTS1/14 animals results in the in-frame skipping of exon 4 (highlighted in Figure 2.4), which would essentially result in a 23 amino acid deletion in the encoded protein. The production of an mRNA transcript at approximately wild-type levels by the mutant allele (best observed in the heterozygous RT-PCR product in Figure 2.4.B) suggests that a protein may be made. The potential protein product of the *Dnajc20* mutation is investigated in chapter 4. This 23 amino acid deletion would not disrupt the J-domain of Dnajc20, which is encoded by exons 2 and 3, but it would remove the beginning of a C-terminal zinc finger present in this protein (Figure 2.5). The identification of this C-terminal domain will be reported in Chapter 3 and its functional relevance discussed in chapter 4. As the MUTS1/14 mutation has been identified, further references will refer to the *Dnajc20* mutation.

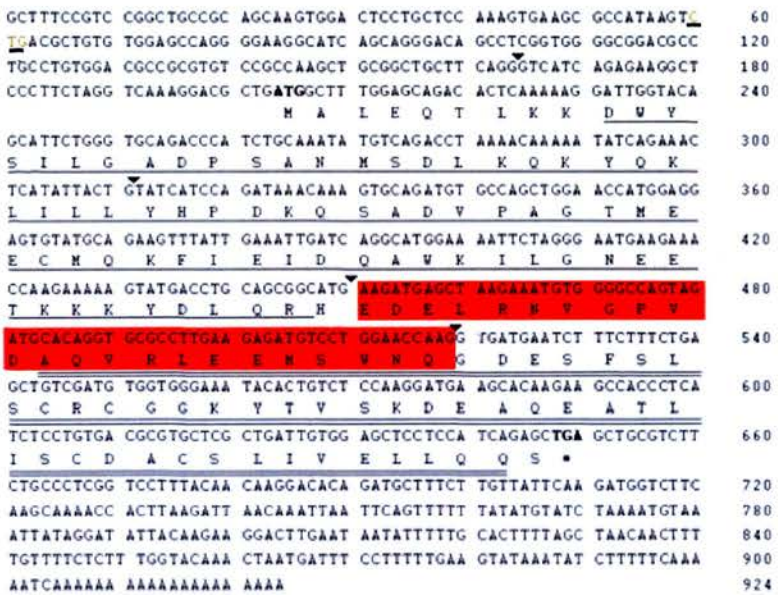


Figure 2.5 The cDNA and protein sequence of *Dnajc20*

The cDNA sequence is from Riken clone 1700030A21. ▼ illustrates the exon boundaries of *Dnajc20*. Exon 4, lost through exon skipping as a result of the *Dnajc20* mutation is highlighted in red. Single line represents the J-domain, the double line represents the C-terminal domain. The underlined CTG was originally published as the start site.

The *Dnajc20* gene has previously been published as *mmDjc7*, a J-protein family member that encodes a type III cytosolic J-protein[22]. The mouse family of J-proteins and their cellular functions is discussed in detail in chapter 3, with particular reference to Dnajc20. The *mmDjc7* gene was reported to encode a 196 amino acid protein. This is in contrast to the 148aa protein predicted by my sequencing of the exons, and the annotation in the various databases. The authors identified an ATG in exon 1, which they suggested as the start site. However, following the sequencing of this exon in a number of different mouse strains, this ATG only appears to be present in the cDNA clone AY028460, used in the characterisation of *mmDjc7*[22]. In the other strains this position is a CTG (thick underline in Figure 2.5) and the published start site for *mmDjc7* is most likely due to a sequencing error.

2.3.2 Phenotype of the *Dnajc20* homozygous mutant embryos

After the *Dnajc20* mutant line was rederived in our facility, a series of crosses on to the C3H genetic background was performed. The C3H background was used as it had been used in the maintenance of the line at MRC Harwell and it has a number of genetic markers that differ from the BALB/c genetic background for genotyping. The BALB/c content of the *Dnajc20* mutant line had been inherited from the mutagenised founder and would therefore contain additional ENU induced mutations. Several generations of crosses on to the C3H background to remove any remaining ENU induced changes that may affect the phenotype of *Dnajc20* mutants.

The *Dnajc20* mutation carriers were followed through the initial crosses through the use of the Mit markers, distinguishing between the BALB/c background in which the mutation was present and the C3H background, which was wild-type for the mutation. However, there appears to be a recombination hotspot immediately distal of the *Dnajc20* gene, which resulted in problematic typing of mutation carriers. Therefore, following the preliminary crosses to C3H, typing performed by sequencing the ENU induced mutation in intron 4 of *Dnajc20*.

Intercrosses between *Dnajc20* heterozygous were performed to better characterise the homozygous phenotype caused by the mutation, and determine the time of prenatal death of the majority of the homozygotes. Contrary to results from MRC Harwell, no homozygous *Dnajc20* mutant mice were born in from these crosses, suggesting that the phenotype observed is more severe on the C3H background than on a mixed C3H/BALB/c background. To investigate the time of lethality in the *Dnajc20* homozygous mutant embryos a series of intercrosses were performed, and embryos recovered at various stages throughout embryogenesis. As a small proportion of the homozygotes had previously been viable it was speculated that embryonic death would most likely occur at later stages of development. At embryonic day 14 and later homozygous embryos were recovered, however, all of these had died at earlier stages and had begun to be resorbed (Figure 2.6.A to C). The mutant embryos at these later stages show gross morphological abnormalities and are much smaller than their wild-type littermates.

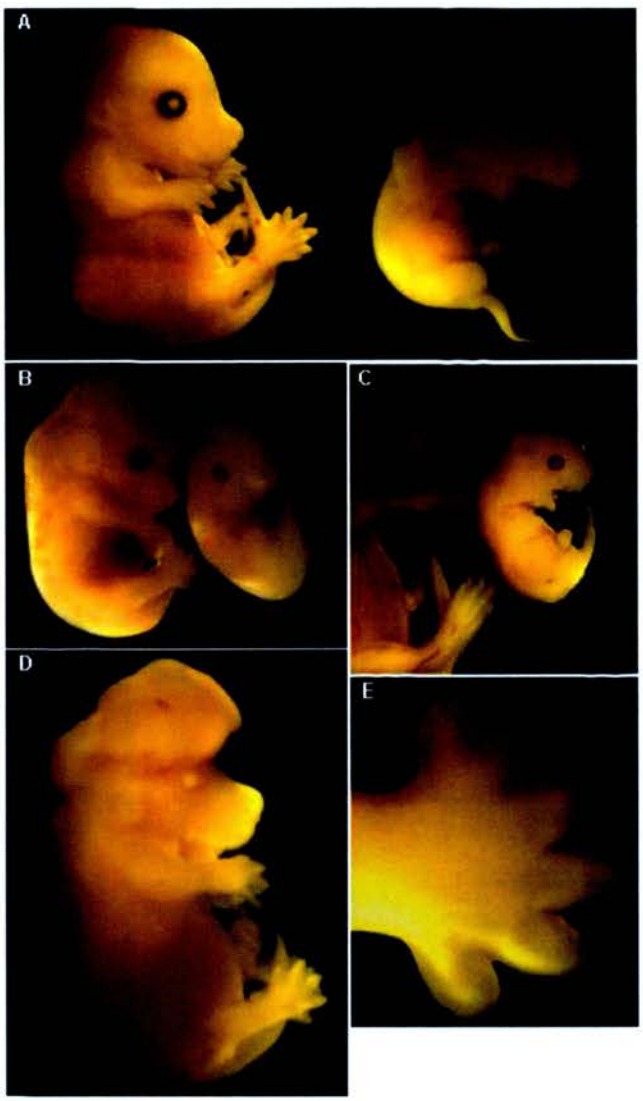


Figure 2.6 *Dnajc20* homozygous embryos

2.6 A to C, homozygous *Dnajc20* mutant embryos (on the right in each image) with wild-type littermates after E14.5. Mutant embryos are developmentally delayed and are partially resorbed. 2.6.D and E. *Dnajc20* homozygous mutant embryo with, 2.6.D, exencephaly and 2.6.E preaxial polydactyly of the right hind limb

The resorption of the embryos makes it difficult to predict the stage when the homozygous embryos were dying, but from the visible morphology, such as the lack of digit formation (Figure 2.6.A to C), the embryos had died (or development had arrested) by embryonic day 13. Phenotyping of the mutant embryos is also made problematic because of resorption. The semi-resorbed embryos are very delicate and therefore, difficult to handle. Furthermore, it is difficult to determine the extent the mutation plays in the phenotype compared to degeneration as a result of resorption.

At E14.5 one mutant embryo was observed with polydactyly of its right hind limb (Figure 2.6.D and E). This is the only time polydactyly was observed in these breeding experiments as no other homozygous mutant embryos reached the stage where they developed digits. This embryo also had exencephaly, a phenotype not observed with any other Dnajc20 homozygous mutant embryos (Figure 2.6.D). Table 2.3 shows the numbers of Dnajc20 mutant embryos recovered at various stages of embryogenesis on the C3H genetic background. The observed numbers of *Dnajc20* mutants implies that death of the embryos starts between E9.5 and E10.5, although a minority do survive until later stages, however overall, the numbers obtained were low at most of the stages. This was predominantly due to small litters on the C3H background. There was a particularly low frequency of *Dnajc20* heterozygous female progeny from crosses to C3H making the study of the homozygous embryos from intercrosses difficult. Attempts to replicate the phenotype seen at MRC Harwell, and increase progeny number, by backcrossing on to BALB/c had no immediate effect.

Table 2.3 *Dnajc20* embryos 1

	Wild-type	Heterozygotes	Homozygotes
E8.5	3	10	3
E9.5	3	4	1
E10.5	10	15	3
E11.5	7	12	2
E12.5	6	12	3
E13.5	10	26	5
E14.5 and above	16	36	9

Table 2.3. On a C3H background *Dnajc20* homozygous mutants begin to die by E10.5. After E13.5 the majority are only present as resorption sites.

In order to reduce the severity of the Dnajc20 homozygous phenotype, as well as increase litter sizes to produce more heterozygous animals, the mice have recently been crossed on to the outbred CD1 background. After one generation of outbreeding, intercrosses were performed between heterozygous littermates and the homozygous phenotype examined from E10.5 to P1. This breeding strategy appears to initially be giving the desired results, an increased number of mutant embryos and better survival. Table 2.4 shows the numbers obtained from these crosses.

Table 2.4 *Dnajc20* embryos 2

	Wild-type	Heterozygotes	Homozygotes
E10.5	2	8	10
E11.5	3	7	3
E12.5	0	6	1
E13.5	8	17	8
E15.5	7	10	2
E16.5	3	6	1

Table 2.4. The genotypes of embryos after breeding to CD1.

The homozygous embryos are easily distinguishable by phenotype is visible from E10.5, the earliest stage examined on this mixed background (Figure 2.7.A). *Dnajc20* homozygous mutant embryos are small throughout embryonic development, are delayed in development (Figures 2.7.A to F) and show polydactyly (Figure 2.7.G). Figure 2.7.C best illustrates the developmental delay. The mutant E13.5 embryo is similar in size and development to a wild-type E12.5 embryo. The developmental delay can also be seen by comparison of the mutant E16.5 embryo (Figure 2.6.F) with the wild-type E15.5 embryo (Figure 2.6.E).

After E13.5 there is a significant fall in the number of expected mutant embryos. Figure 2.7.D shows a wild-type E13.5 embryo with two mutant littermates and provides evidence for the varying severity of the *Dnajc20* mutant phenotype. Both mutant embryo are small (being similar in size to an E12.5 embryo, Figure 2.7.C) and developmentally delayed. However the extent of development varies enormously, one of the mutant embryos (centre in Figure 2.7.D) appears to be delayed by nearly 24 hours in development, whereas the other mutant embryo (right in Figure 2.7.D) is delayed by at least 48 hours in the development of the limb bud. In this embryo development seems to have arrested, but as it is a similar size to its mutant littermate, growth appears to have continued. It is possible to speculate that mutant embryos that are delayed by up to 24 hours may be able to survive until later stages of development, whereas the embryos where development appears to have arrested will be lost. *Dnajc20* homozygous mutant embryos are able to survive further in development on this genetic background, and will allow more detailed study of their embryonic postnatal phenotype. Out of 20 pups born from 2 litters one *Dnajc20* mutant has been born (Figure 2.8.A). This P1 animal was smaller than its wild-type and heterozygous littermates (mutant weight was 0.9g compared to an average wild-type weight of 1.85g and heterozygote weight of 1.81g). The mutant animal didn't feed but was alive at birth. Figure2.8.B to D shows the small size of the mutant internal organs compared to those of a heterozygous littermate. Figure 2.8.E illustrates delayed development of the mutant spleen.

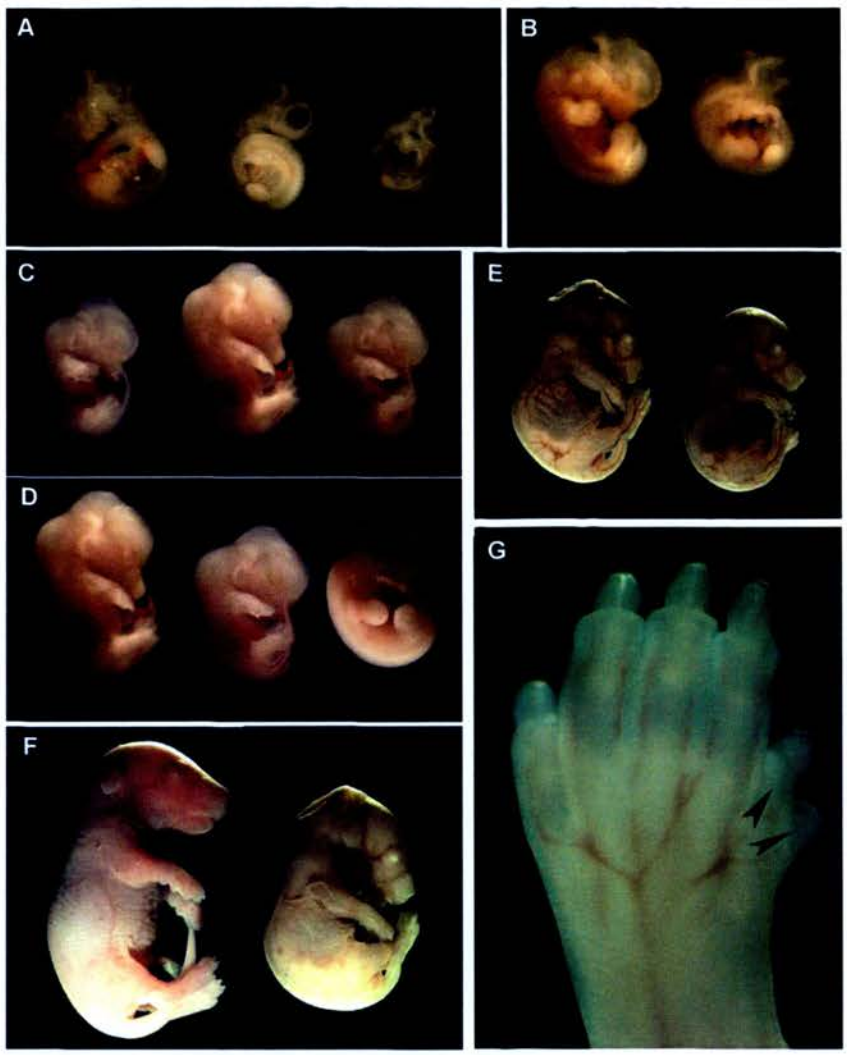


Figure 2.7 *Dnajc20* homzygous mutant phenotype

2.7.A to F. *Dnajc20* homozygous mutant embryos through development. 2.7.A. E10.5 wild-type embryo (*left*)and two mutant E10.5 embryos (*right*). 2.7.B. E11.5 wild-type embryo (*left*) and E11.5 mutant embryo (*right*). 2.7.C. E12.5 wild-type embryo (*left*), E13.5 wild-type embryo (*centre*), and E13.5 mutant embryo (*right*) illustrating the developmental delay of the mutant embryo. 2.7.D. E13.5 wild-type embryo (*left*), ad two E13.5 mutant embryos (*right*). The embryo on the right displays severe developmental delay, but is of similar size to the other mutant embryo. 2.7.E. E15.5 wild-type embryo (*left*) and mutant E15.5 embryo (*right*). 2.7.F. Wild type E16.5 embryo (*left*) and mutant E16.5 embryo (*right*). 2.7.G. Left hind limb of E16.5 mutant embryo in 2.7.F showing the preaxial polydactyly (*arrows*).

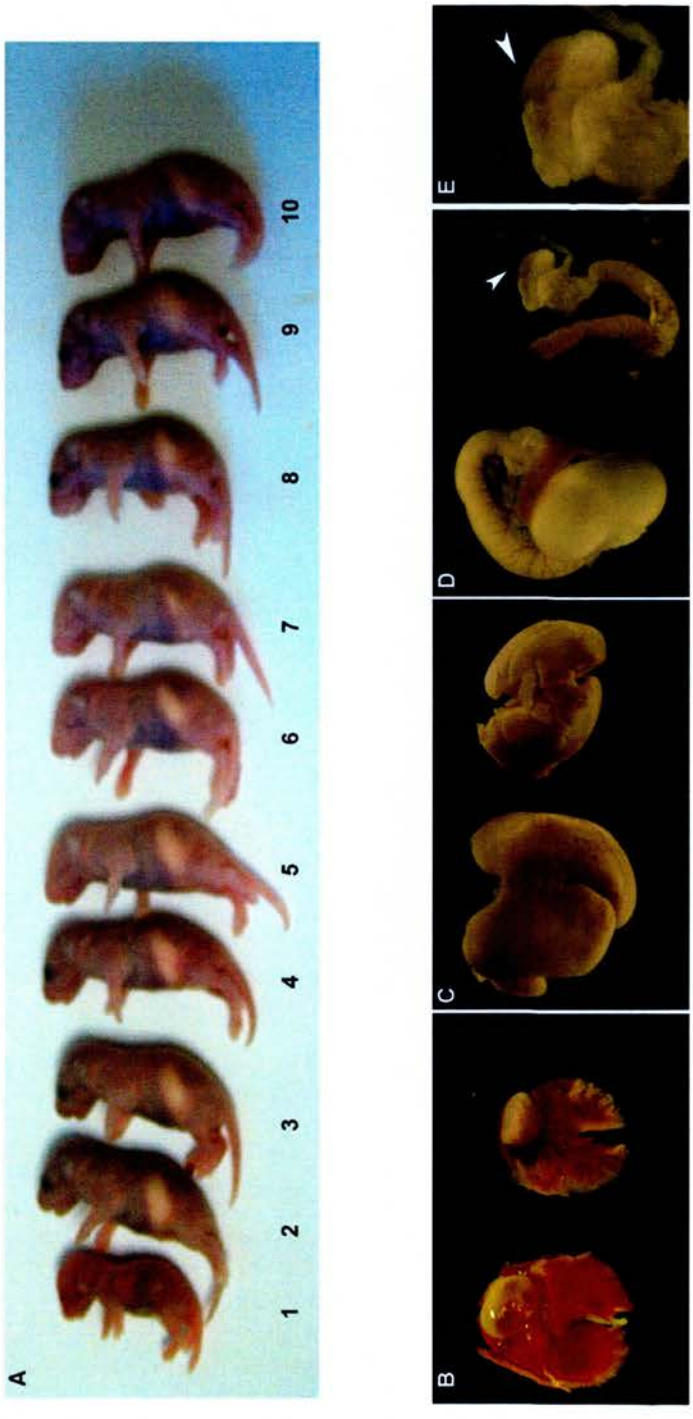


Figure 2.8 *Dnajc20* P1 homozygous mutant mouse

2.8.A. From left to right (1)P1 mutant mouse,(2) Heterozygote (Het), (3) Het, (4) Wild-type (WT), (5) WT, (6) Het, (7) WT, (8) Het, (9), Het, (10) WT. 2.8.B to E. Illustrate small internal organ size of mutant (1) P1 mouse. Heterozygote (2) P1 mouse was also dissected. 2.8 .B Heart and liver (mutant on right). 2.8 .C Liver (mutant on right). 2.8.D Stomach and spleen (mutant on right). 2.8.E. Magnified mutant spleen. Mutant mouse has an immature spleen phenotype (*arrows*).

2.4 Discussion

A large number of mutagenesis screens have been performed in the mouse with the goal of determining gene function via the effects of mutation. Our group, in collaboration with MRC Harwell, has recently performed an ENU screen targeted against a 3Mb region around the *Pax6* gene on mouse chromosome 2. One mutation has been identified from this screen and results, predominantly, in embryonic death around midgestation. The homozygous embryos display growth retardation and are developmentally delayed, with some having preaxial polydactyly on one or both hind limbs. By sequencing the genes present in the candidate region I have identified a mutation in a gene encoding a J-protein family member of unknown function. The mutation, a T to A transversion at position +2 of the splice donor site in intron 4 of *Dnajc20*, results in the in-frame skipping of exon 4, essentially causing a 23 amino acid deletion of the potential protein.

These findings reveal that, on the whole, *Dnajc20* is necessary for embryonic and postnatal survival. The primary phenotype of the *Dnajc20* mutant homozygotes is their universal reduction in size, as well as preaxial polydactyly in the majority of cases. Only three *Dnajc20* mutant homozygotes have survived past weaning, and these were all from the initial intercrosses performed at MRC Harwell. These were small compared to wild-type and heterozygote littermates and had polydactyly on their hindlimbs. However, only one homozygous animal has been born since. On a C3H background all

of the homozygous embryos appear to die by E13.5, although some embryos can be recovered from resorption sites at later stages. On this background there is also a reduction in the expected number of mutant embryos from around E10.5. The presence of one generation of CD1 crosses to the phenotype seems to rescue a fraction of the mutants. The majority still appear to die by E13.5, but there is now a subpopulation that can survive until after birth. An interesting feature of the phenotype is shown in Figure 2.6.D. Here, two E13.5 mutant littermates illustrate the two phenotypic subclasses caused by the mutation in *Dnajc20*. The two mutant embryos are of a similar size, however, there is obvious developmental delay in one of them. This embryo was removed from a resorption site and would represent the population that die by embryonic day E13, whereas the other mutant from this litter may be able to survive until birth. Although the numbers are small, especially in the case of the recent results obtained by crossing to the CD1 mice, they do suggest the presence of genetic differences between the genetic backgrounds that modify the *Dnajc20* mutant phenotype. Presence of the CD1 into the crosses will allow the more detailed characterisation of the *Dnajc20* homozygous phenotype, and the dissection of the physiological role of the Dnajc20 protein.

An obstacle in the use of ENU mutagenesis is also one of its benefits. As ENU induces a mutation at a rate of approximately 1 per megabase[31], there is the possibility that any base change identified may not be the mutation responsible for the phenotype. In order to be sure that the identified mutation is indeed the cause of the phenotype further

investigation may be needed. Techniques used include BAC rescue, whereby the wild-type allele of the candidate gene is reintroduced to the mutant mouse to rescue the phenotype. Having a second mutant allele in the candidate gene also allows confirmation of the candidate. In the case of *Dnajc20* there are a number of gene-trap cell lines that disrupt the gene. We have imported one of these gene-trap lines, an insertion in intron three that will result in the loss of the final two exons, and this will provide definitive proof that the mutation in *Dnajc20* does result in the observed phenotype. However, even without the gene-trap mutation, there are a number of convincing factors that the identified *Dnajc20* mutation is responsible for the phenotype. The strategy of the screen and subsequent recombination localises the mutation to a 1.4Mb region. The size of the region suggests the slight possibility of an additional ENU induced change. However, the majority of the coding sequence of the region has been sequenced and the only change identified was that in *Dnajc20*. The most striking evidence in the verification of this mutation, however, is the phenotypic similarity to a mutation in another gene, which is involved in the same biological process as *Dnajc20* and is discussed in chapter 4.

The use of ENU mutagenesis to provide mutations in order to define protein function is proving highly successful. Here, I have identified a mutation in a J-protein of unknown function. The challenge now is to identify the role of Dnajc20 in a biological context. As a starting point a sequence-based approach has been used and this is reported in Chapter 3.

Chapter 3

Dnajc20 and the mouse family of J-proteins

3.1 Introduction

The identification of a mutation in Dnajc20 resulted in an interest in this large family of molecular chaperones. However, upon initial investigation it became apparent that the annotation of the J-protein family has been problematic, with numerous omissions and misannotations. A number of different nomenclatures are commonly used in describing members of this family and this is further complicated by the some genes having multiple names. In this chapter I have used a bioinformatics approach in the classification of the mouse J-protein family as a basis for the further study of Dnajc20.

3.1.1 The Hsp70-chaperone machinery

The term molecular chaperone describes a wide group of unrelated protein families that are principally responsible for preserving a functional set of cellular proteins. The majority of molecular chaperone families are called heat shock proteins (Hsps) as their initial discovery was due to their rapid induction following stress. The Hsp70-chaperone machinery is one of the most studied examples of molecular chaperones and is composed of the Hsp70 proteins themselves and their co-chaperones the J-proteins. These chaperone pairs form an abundant set of machinery that is highly conserved throughout evolution and present in all biological kingdoms[53]. Hsp70-chaperone machines are involved in many processes in cell survival and normal cell physiology including; 1) folding of newly synthesised proteins, 2) translocation of proteins across

membranes, 3) transportation of proteins to their correct location, 4) assisting the formation of protein complexes, 5) targeting proteins for degradation. In addition to these classical roles of the Hsp70-chaperone machinery in protein folding, more specialised regulatory roles are emerging in signalling pathways, and the maintenance of intermediate states of protein substrates[54].

3.1.2 The Hsp70 proteins

Hsp70 proteins are highly conserved and contain the same overall structure throughout evolution. There are at least 14 Hsp70 proteins in eukaryotes[55], and the conservation between mouse Hspa1b, previously named Hsp70, and *E.coli* DnaK, the characteristic member of the Hsp70 protein family, is 50% identity at the protein level[56]. They are made up of three domains, an N-terminal actin-like ATPase domain, the substrate binding domain (SBD), and a C-terminal domain involved in the interaction with J-proteins.

The mechanism of substrate recognition by the SBD is relatively non-specific. The SBD recognises short hydrophobic segments within the substrate polypeptide, which comprises 5 amino acids, flanked by positively charged residues[57]. It is less clear, however, how Hsp70 proteins bind folded substrate proteins. To be identified as a substrate a folded protein might have to contain an exposed structure, either by locally unfolding or by containing an intrinsically unfolded hydrophobic segment[58].

3.1.3 The Hsp70 ATPase cycle

The activity of Hsp70 proteins is reliant upon the binding and release of ATP, which in turn regulates the activity of the Hsp70 proteins interaction with its substrate via the SBD[59]. Binding of ATP to the N-terminal ATPase domain of Hsp70 decreases the affinity of the SBD with the substrate. Consequently, Hsp70 protein cycles between an ATP bound state where the SBD is open, and an ADP state where the SBD is closed and the substrate is trapped[57,60] (Figure 3.1). The affinity of Hsp70 alone for ATP is minimal and is strongly increased by J-proteins[61,62]. It is now accepted that the specificity of the Hsp70-chaperone machinery is provided by the J-proteins, whereby presentation of a substrate protein is coupled to the stimulation of ATPase activity of its Hsp70 partner[54,63].

3.1.4 J-proteins

The name J-protein originates from the *E.coli* protein DnaJ, the prototypical member of the family. The *DnaJ* gene of *E.coli* was originally identified because when mutated it blocked bacteriophage λ DNA replication[64-66]. Other mutations were shown to be detrimental to bacterial growth at high temperature, providing the name for the gene[67]. The *E.coli DnaJ* gene forms an operon with the *DnaK* gene, the bacterial homologue of the eukaryotic Hsp70s[65].

The 37kDa DnaJ protein of *E.coli* is composed of four domains (Figure 3.2.A), and is the best studied of the J-proteins. At the N-terminus of DnaJ is the J-domain, a structure consisiting of four α-helices that is required for the stimulation of its Hsp70 protein partner, DnaK.

Helices II and III form an antiparallel strucure seperated by an exposed loop which contains the tripeptide Histidine-Proline-Aspartic acid (HPD) motif, all three residues of which are essential for the stimulation of DnaK. Helices I and IV at either end of the J-domain are less well conserved and function in stabilising the antiparallel structure of the central helices. Adjacent to the J-domain is a region of approximately 40 amino acids, which is rich in glycine and phenylalanine (GF) residues. This structure is thought to act as a flexible linker region[56], and is required for the maximum stimulation of Hsp70 proteins by the J-domain[68]. Distal of these two domains is a cysteine rich zinc-binding domain, which is involved in the binding of unfolded protein substrates and as an interaction site with Hsp70 proteins[69]. These three domains are contained in the first half of the protein, the second half contains the C-terminal domain (CTD) which is poorly understood and less well conserved between different J-proteins [56].

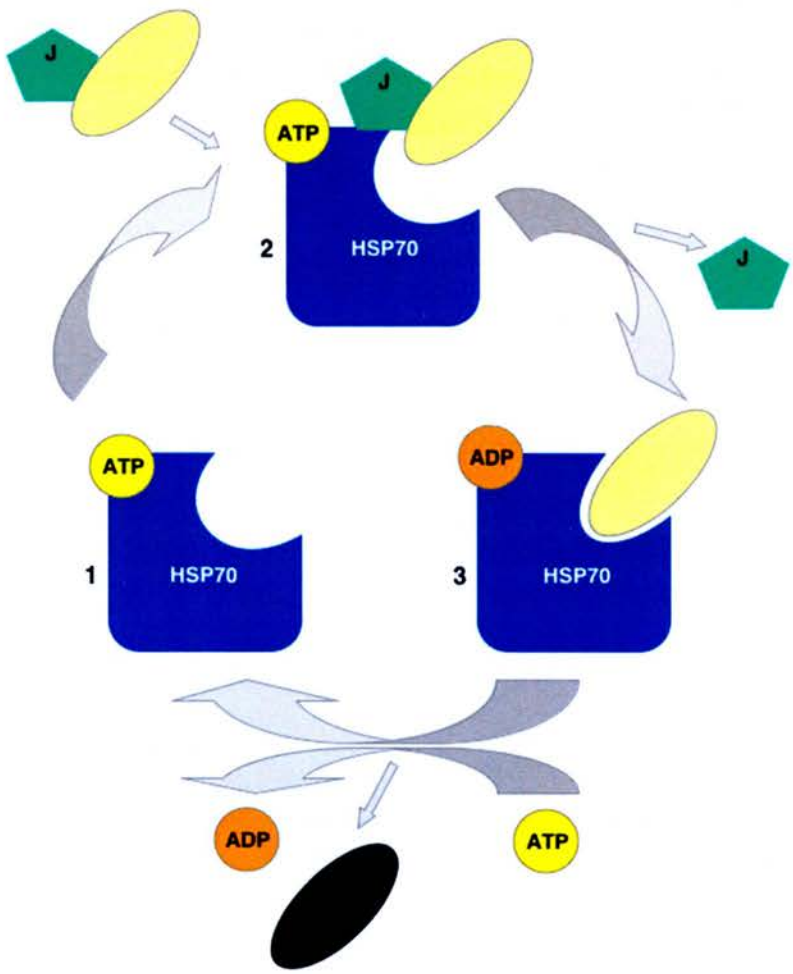


Figure 3.1 The Hsp70 ATPase cycle

The regulation of Hsp70 ATPase activity and substrate binding is controlled by the co-chaperone J-protein. 1. Hsp70-ATP has low affinity for the substrate protein. 2. The J-protein presents a substrate protein to its Hsp70 partner and stimulates its ATPase activity resulting in 3. Hsp70-ADP with high affinity for the substrate protein. The modified substrate protein is released, and Hsp70 returns to an ATP bound state.

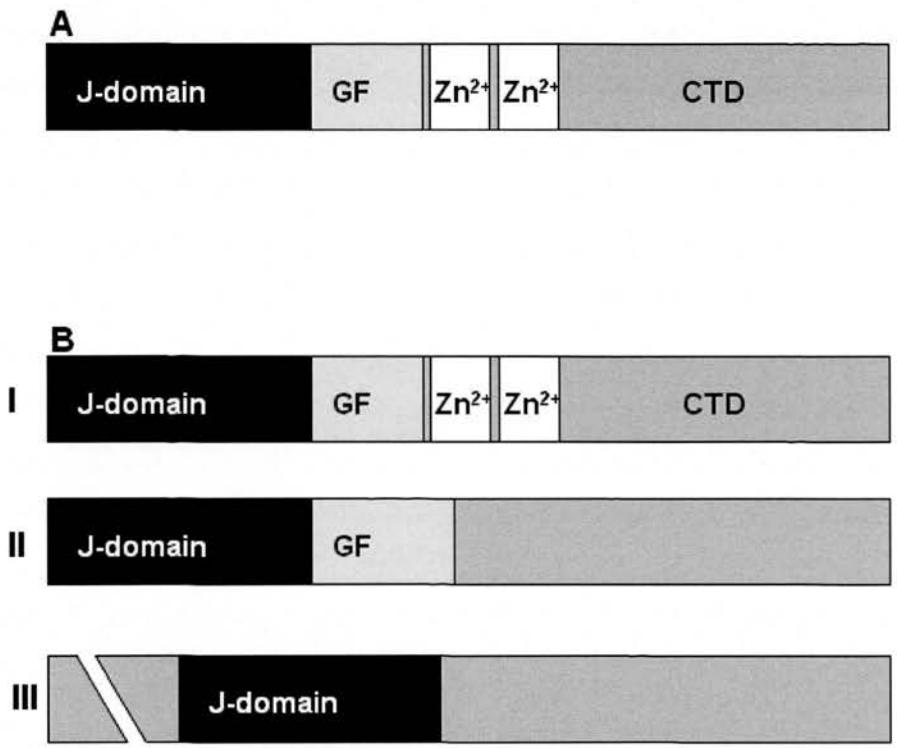


Figure 3.2 *E.coli* Dnaj and the J-protein subclasses

Adapted from Cheetham and Caplan, Cell Stress & Chaperones (1998)[56].

A. The domain structure of *E.coli* DnaJ, an N-terminal J-domain, GF rich region, Cysteine rich zinc binding domain, and the poorly conserved C-terminal domain (CTD).

B. Sub-classification of the J-proteins. Type I J-proteins share full domain structure with *E.coli* DnaJ. Type II J-proteins contain an N-terminal J-domain and the GF rich region. Type III J-proteins only have a Jdomain in common with *E.coli* DnaJ, and this can be anywhere within the protein.

3.1.5 The nomenclature of the J-proteins

The J-protein family has previously been described by a number of terms including the Hsp40 proteins, DnaJ homologues, J-domain proteins, and a variety of combinations of these names. The term J-proteins has recently been coined to describe the large family of molecular chaperones that contain a J-domain, defined by the presence of four helices and the HPD motif[53]. This definition of the family also distinguishes the J-proteins from the J-like proteins, which contain a domain similar to the J-domain but lack the HPD motif. The J-like proteins, such as Pam16 in yeast[70], are unable to activate the ATPase activity of HSP70 proteins. This definition of a J-protein follows the establishment of a sequence based classification of the J-protein family[56] and a naming system that reflects this has been suggested[71]. An evolved form is now in place in the higher eukaryotes. The nomenclature of the J-protein family had previously been very complicated with many genes having multiple names under different classification systems. Due to this, the J-proteins discussed here will be referred to using the current J-protein nomenclature system as well as by their other common names.

3.1.6 Sub-classification of the J-proteins

The J-protein family encompasses a diverse group of proteins with very different domain structures, the only common feature being the J-domain. This variety led to the classification of the J-proteins into three sub-groups based upon their overall

conservation with *E.coli* DnaJ[56] (Figure 3.2.A and B). In this classification system the J-proteins, which have full domain conservation with DnaJ, are called the type I J-proteins. Type II proteins have an N-terminal J-domain as well as the GF rich region, and the type III J-proteins only have the J-domain in common with DnaJ, the position of which varies within the protein.

It has been suggested that some J-proteins may have acquired the J-domain to present a particular substrate to an Hsp70 partner, and utilise the chaperone activities for a specific role instead of a more general chaperone function[54]. This type of specialised chaperone is thought to be made up predominantly of type III J-proteins, with the type I and type II J-proteins performing a more general chaperone function[53,54]. By definition, all of the type I J-proteins contain the same domain complement as *E.coli* DnaJ, and are likely to be functionally equivalent. For example, the mammalian type I J-protein Dnaja1 (Hsj2/Hdj2), a partner of the constitutively expressed cytosolic Hsp70 protein Hsc70, is required for the refolding of denatured proteins and protein import into the mitochondria[72,73]. In addition to their J-domains and GF rich regions some of the type II J-proteins contain additional domains, although these are also linked to a traditional chaperone function, such as the ubiquitin interaction motif of the neuronally expressed Dnajb10 (Hsj1). This J-protein functions in the prevention of protein aggregation and stimulates the sorting of proteins to the proteosome[74-76].

One of the best understood examples of a type III J-protein system is the clathrin coat disassembly mediated by bovine auxilin (Dnajc6)[77-80]. Auxilin is a 910 residue protein composed of an N-terminal tensin-like domain, a central clathrin binding domain, and a C-terminal J-domain. Auxilin binds to clathrin-coated vesicles via its clathrin binding domain and then recruits its Hsp70 partner, Hsc70 through its J-domain. In a reaction coupled to ATP-hydrolysis, auxilin and Hsc70 disassemble the clathrin coat to form a transport vesicle. Interestingly other J-domain proteins have also been identified in this process, including Cyclin G-associated kinase (GAK), which, outside of its kinase domain shares high homology with auxilin[81-83]. The *Drosophila* orthologue of mammalian Dnajc13, Rme8 has also been suggested to be involved in a clathrin dependent process[84].

3.1.7 Partner preference

The number of J-proteins that have been identified to date is much greater than the number of Hsp70 proteins. For example there are 22 J-proteins in yeast compared to only 14 Hsp70 proteins[53]. This difference in numbers is explained by Hsp70 proteins being able to interact with multiple J-proteins. A good example is the Hsp70 partner of auxilin Hsc70. This Hsp70 protein is the partner of the other J-proteins involved in clathrin dependent processes, as well as Dnajc10 (Hsj1)[74], Dnaja1 (Hdj2, Hsj2)[73], and Dnajc5 (Csp)[85]. Not all Hsp70 proteins have multiple J-protein partners, for example, the yeast Hsp70 Ssz1 only interacts with the J-protein Zuo1[86].

The J-proteins have thus far only been shown to have one *in vivo* Hsp70 partner each. Domain swapping experiments and *in vitro* studies have shown that some J-domains are capable of stimulating the ATPase activity of multiple Hsp70 proteins, for example *E.coli* DnaJ is capable of stimulating mammalian Hsc70[87]. Dnajc20 was published as a J-protein that is able to stimulate the ATPase activity of Hsc70, the endoplasmic reticulum resident Hsp70 BiP, and *E.coli* DnaK[22]. Experiments such as these suggest that some J-proteins have the potential to stimulate multiple Hsp70 proteins, however, additional factors such as their localisation will ultimately determine whether these interactions occur under normal physiological conditions.

Over the past few years it has been established that the functional specificity of the Hsp70-chaperone machinery is provided by the J-proteins, with the structurally diverse type III J-proteins in particular being involved in a wide range of different functions. The hypothesis that the J-domain of these type III J-proteins has been recruited to bring chaperone power to a particular process, the other structural features of these proteins may offer important clues in assigning a function.

3.1.8 Mutations in the mouse J-proteins

The analysis of mutations within the J-protein family has provided a wealth of information regarding their function, and the requirements of the J-domain. The

majority of this work was performed in *E.coli* DnaJ and, as stated previously, first identified the J-proteins as molecular chaperones[65]. The systematic deletion of yeast ORFs also means there is a ready made knock out of each of the yeast J-proteins, and this resource has been used in their characterisation since the first yeast DnaJ homologue was identified[88]. Much less is known about the effect of mutations in the J-protein family in higher eukaryotes. Dnajc20 is only the sixth mouse J-protein which has been mutated, and the first by ENU mutagenesis.

The first of the mouse J-proteins to be mutated in the mouse was the type II Dnajb6 (Mrj), which was identified in a gene trap screen for developmentally important genes (Mrj)[89,90]. Homozygous mutants die at E18.5 due to a failure of chorioallantoic fusion. Dnajb6 (Mrj) was shown to be expressed in the placenta and developing embryo, as well as in some adult tissues, the early lethality precluded study of its role in these tissues. The finding that it was required for formation of the chorioallantic placenta showed for the first time the requirement for some J-proteins in developmental processes.

Over the past year, four more mouse J-proteins have been mutated by targetted knock out. Mutation of the type I J-protein Dnaja1 (DjA1, Hsj2, Hdj2) resulted in defects in the late stages of spermatogenesis[91]. Dnaja1 (DjA1, Hsj2, Hdj2) is widely expressed in the mouse, but is potentially functionally redundant with Dnaja2 (DjA2, mDj3), which has similar expression. *Dnaja3* (*mTid-1*) is the orthologue of the *Drosophila l(2)tid* gene,

Tid56, the first J-protein to be identified as a tumour supressor[92]. *Dnaja3*$^{-/-}$ embryos die between E4.5 and E7.5, due to unknown defects[93].

Two type III J-proteins have also been knocked out recently, Dnajc5 (CSP), an abundant synaptic vesicle protein[94], and Dnajc3 (p58^{IPK}), which functions in the ER stress response[95]. *Dnajc5*$^{-/-}$ animals are physiologically normal for the first two weeks of life, however, after this time they stop gaining weight, and after P21 they begin to die due to a progressive neurological disorder[94]. The neurological disorder is a result of a loss of integrity at the neuro muscular junctions (NMJ). Mutation of the *Dnajc3* gene results in low body weight due to a decrease in the number of insulin producing β-cells in the pancreatic islets[95]. The Dnajc3 protein is thought to function as an "off switch" to signal the downregulation of ER stress proteins following their initial stress response. The absence of the Dnajc3 protein, therefore results in the misexpression of the ER stress proteins causing an increase in apoptosis in the β-cells.

The five mutations reported in mouse J-proteins result in a variety of phenotypes and provide evidence of the functional diversity of the family in the higher eukaryotes, with essential roles in development, spermatogenesis, neuroprotection, and diabetes.

3.1.9 Dnajc20 and the mouse family of J-proteins

The J-proteins have been most intensively studied in prokaryotes, with the majority of the functional information arising from *E.coli* DnaJ. This family of proteins is much larger in the eukaryotes and the number of J-proteins identified is still increasing. Recently, via the analysis of the *Saccharomyces cerevisiae* genome, Walsh *et al* identified the full complement of J-proteins in a eukaryote revealing the presence of five type I, 4 type II, and 13 type III J-proteins[53].

A number of the individual mouse J-proteins have been well studied including Dnaja1 (Hsj2, Hdj2)[72], Dnajb10 (Hsj1)[96], Dnajc5 (Csp)[97], and Dnajc3 (p58IPK)[95]. However, as a family they remain complicated with a number of omissions and misannotations. Previous studies have revealed the presence of at least 23 J-proteins in the mouse, however, the total number has yet to be determined[71]. The MGI database (http://www.informatics.jax.org) currently lists 39 J-proteins under the accepted nomenclature, whereas, the Ensembl database (http://www.ensembl.org) suggests the presence of 51 J-proteins, based upon the presence of an INTERPRO J-domain. The Dnajc20 protein is identifiable in the Ensembl database by the presence of the INTERPRO definition of the J-domain, but is absent from the list of J-proteins in the MGI database. In order to properly relate the Dnajc20 protein to the other J-proteins, the starting point has to be the correct organisation of the mouse J-protein family.

3.2 Results

3.2.1 Identifying the J-protein family members in mouse

Due to the complicated appearance of the J-protein family in mouse, I decided to use a bioinformatic approach to clarify the situation and potentially confirm the presence of novel J-proteins. The genuine number of J-proteins in the mouse is difficult to gauge, the MGI database currently lists 39 J-proteins under the official nomenclature, and in Ensembl, the Heat shock protein DnaJ N-terminal INTERPRO domain (IPR001623) is annotated as being present at 51 loci (http://www.ensembl.org/Mus_musculus/domainview?domainentry=IPR001623). However, neither of these two databases reflect the true number J-proteins in mouse. The MGI database is likely to contain all of the J-proteins but if they are not named according to the current nomenclature, they will be omitted from the family. In a number of cases, the presence of a J-domain in Ensembl does not correspond to a functional ORF. Additionally, the INTERPRO annotation of a J-domain is not very stringent, and therefore includes a number of proteins that would fall into the category of J-like proteins. However, by using a combination of the information in the MGI database, the presence of the IPR001623 entry in Ensembl, and information in the literature, I have been able to construct a list of genuine J-proteins encoded by the mouse genome as defined by the presence of a J-domain. The J-domain was confirmed by SMART analysis (http://smart.embl-heidelberg.de), as defined by a search with the Pfam HMMER model of the domain, and

its postion within each protein is noted within the tables. Additional domains, identified by SMART analysis or from information in the literature were also recorded, providing an insight into the diverse variety of structural features present within the J-protein family in mouse.

This analysis has revealed the presence of 48 genuine J-proteins encoded by the mouse genome, 4 of which are type I (Table 1), 12 are type II (Table 2), and 32 are type III (Table 3). This may not represent the complete number of J-proteins as the presence of additional potentially expressed sequences, that could encode J-domains within the mouse genome leaves the possibility of further J-proteins. For example, the recently published Gγg10 protein of humans has alternative 5 prime exon which encodes a J-domain[98]. I could not identify an orthologue of this exon in mouse, however, the alternative splicing of a J-domain encoding sequence onto another transcript may occur elsewhere. Under these circumstances the annotation may be poor with the correct transcripts not being annotated in Ensembl, especially if they belong to a poorly expressed splice variants. To confirm the true number of J-proteins in the mouse further work would have to be undertaken in order to confirm the expression of all of the potentially encoded J-domains.

3.2.2 The Mouse J-proteins

Table 1. Type I J-proteins

Gene (alternative)	Ensembl Gene ID	Length (J-domain)	Other Domains
Dnaja1 (Hsj2, Nedd7)	ENSMUSG00000028410	397 (6-68)	None
Dnaja2 (mDj3)	ENSMUSG00000031701	412 (8-70)	None
Dnaja3 (mTid-1)	ENSMUSG00000004069	480 (93-158)	None
Dnaja4 (mmDjA4, Hsj4)	ENSMUSG00000032285	397 (6-68)	None

Table 2. Type II J-proteins

Gene (alternative)	Ensembl Gene ID	Length (J-domain)	Other Domains
Dnajb1 (Hsp40, Hdj1)	ENSMUSG00000005483	340 (4-68)	None
Dnajb3 (Hsj3, Msj1)	ENSMUSG00000048591	242 (3-69)	None
Dnajb4	ENSMUSG00000028035	337 (4-68)	None
Dnajb5 (Hsc40)	ENSMUSG00000036052	348 (4-68)	None
Dnajb6 (Mrj)	ENSMUSG00000029131	242 (3-69)	None
Dnajb7 (mDj5)	ENSMUSG00000047108	220 (2-61)	None
Dnajb8 (mDj6)	ENSMUSG00000048206	227 (3-69)	None
Dnajb9 (Erdj4, Mdg1, mDj7)	ENSMUSG00000014905	222 (26-90)	Transmembrane
Dnajb10 (Hsj1,mDj8)	ENSMUSG00000026203	259 (3-69)	UIM (Ubiquitin interaction motif)
Dnajb11 (Erdj3)	ENSMUSG00000004460	358 (25-90)	Transmembrane
Dnajb12	ENSMUSG00000020109	376 (111-175)	Transmembrane
Dnajb14	ENSMUSG00000059332	379 (107-164)	TPR (Tetratricopeptide repeat), N-6-Adenine Specific Methyltransferase

Table 3. Type III J-proteins

Gene (alternative)	Ensembl Gene ID	Length (J-domain)	Other Domains
Dnajc1 (Erdj1, Mtj1, Dnajl1)	ENSMUSG00000026740	552 (61-125)	Transmembrane, SANT domain
Dnajc2 (Mida1, Zrf1, Zrf2)	ENSMUSG00000029014	621 (88-161)	SANT domain
Dnajc3 ($p58^{IPK}$)	ENSMUSG00000022136	504 (394-462)	Tetratricopeptide Repeat
Dnajc4 (Hspf, Mcg18)	ENSMUSG00000024963	244 (37-102)	Transmembrane
Dnajc5 (Csp)	ENSMUSG00000000826	198 (15-80)	Cysteine string (CS)
Dnajc5β	ENSMUSG00000027606	199 (19-84)	Transmembrane (CS)
Dnajc5γ	ENSMUSG00000053856	165 (17-82)	CS
Dnajc6 (auxilin)	ENSMUSG00000028528	938 (874-838)	PTEN clathrin binding
Dnajc7 (mDj11,mTpr2, Ttc2)	ENSMUSG00000014195	494 (381-451)	Tetratricopeptide Repeat
Dnajc8	ENSMUSG00000054405	292 (57-123)	Bipartite nuclear localisation signal
Dnajc9	ENSMUSG00000021811	259 (15-82)	None
Dnajc10 (Jpd1)	ENSMUSG00000027006	793 (35-100)	Thioredoxin
Dnajc11	ENSMUSG00000039768	559 (14-82)	None
Dnajc12 (Jdp1)	ENSMUSG00000036764	198 (14-79)	None
Dnajc13 (Rme8)	ENSMUSG00000032560	2152 (1210-1274)	tRNA synthetase
Dnajc14 (Drip78, Hdj3, Lip6)	ENSMUSG00000025354	703 (444-508)	Bipartite nuclear localisation signal
Dnajc15 (Dnajd1)	ENSMUSG00000022013	149 (94-147)	None
Dnajc16	ENSMUSG00000040697	820 (77-141)	None
Dnajc17	ENSMUSG00000034278	303 (11-76)	RNA Recognition Motif
Dnajc18	ENSMUSG00000024350	357 (82-146)	None
Dnajc19 (Dnajd2)	ENSMUSG00000057330	116 (62-115)	Transmembrane
Dnajc20 (1700030A21Rik, mmdjc7)	ENSMUSG00000027166	148 (10-81)	CSL zinc finger
4930461P20Rik (Dnaja5)	ENSMUSG00000044224	531 (3-69)	C2H2 zinc finger
GAK (auxilin2)	ENSMUSG00000062234	1,305 (1231-1305)	Protein Kinase, PTEN clathrin binding
Hsc20_mouse (Hscb)	ENSMUSG00000043510	234 (155-229)	None
Sec63-like	ENSMUSG00000019802	760 (104-165)	Transmembrane
RBJ	ENSMUSG00000020657	273 (217-273	RAS-transforming
Sacsin	ENSMUSG00000037545	3830 (3557-3644)	Heparin, N-terminal Hsp90 like
Wbscr18	ENSMUSG00000063758	219 (42-107)	None
4930503B20Rik	ENSMUSG00000048652	234 (3-69)	None
2819451A06Rik	ENSMUSG00000038009	339 (277-339)	Transmembrane
ORF28	ENSMUSG00000039763	385 (48-113)	None

Elsewhere in this analysis certain misannotations became clear. The Dnajb13 protein (MGI:1916637) has been omitted as it does not contain the HPD motif, and would therefore presumably be unable to stimulate the ATPase activity of an HSP70 protein. There is also no Dnajb2 protein in mouse, with the orthologue of human Dnajb2 being mouse Dnajb10 (Hsj1), and conversely, there is no Dnajb10 in humans. The 4930461P20Rik is the orthologue of the recently published human J-protein Dnaja5, however, this protein does not contain a GF repeat and should be included in the type III family. 4930461P20Rik is a 531 amino acid protein with an N-terminal J-domain and a C-terminal C2H2 zinc finger. The authors probably classified the human Dnaja5 as a type I protein due to the presence of this zinc finger domain, but it is not homologous to the cysteine rich zinc finger of *E.coli* DnaJ[99]. A BLAST search with this protein reveals it to be the sequence orthologue of the type III yeast protein JJJ1[53].

3.2.2 Bioinformatic characterisation of novel J-proteins

All of the type I and type II J-proteins have been named under the proper nomenclature, however, 11 of the type III proteins are either unnamed or named under other criteria. Four of these J-proteins, Wbscr18, 4930503B20Rik, 2819451A06Rik, and ORF28, are novel J-proteins, although, Wbscr18 and ORF28 have previously been identified in large scale investigations into the content of the Williams-Beuren syndrome candidate region in humans[100], and the mouse orthologues of human chromosome 21 genes[101]

respectively. It was decided to investigate these novel J-proteins further through initial bioinformatic characterisation.

4930503B20Rik

The 234 amino acid 4930503B20Rik protein has an N-terminal J-domain, but the remaining sequence shows no homology to any other domains. It has been identified as a transcript that is preferentially expressed in mammalian testis by SAGE analysis. It has orthologues by sequence in mammals, however, there is no annotated human orthologue in Ensembl. There is a human EST from testis, AY604569, that maps to the syntenic region. 4930503B20Rik is annotated as being a mitochondrial protein, however, this annotation appears to be due to its homology to yeast Ydj1. PSORT II analysis (http://psort.nibb.ac.jp/) predicts a nuclear localisation.

2810451A06Rik

2810451A06Rik is a 339 amino acid protein with a C-terminal J-domain, SMART analysis predicts 4 transmembrane domains and an N-terminal signal peptide. PSORT II analysis predicts three of these transmembrane regions with high confidence and the fourth with low confidence as well as a signal peptide, with a potential cleavage site between residues 44 and 45. It is predicted to be an ER resident protein. The

2810451A06Rik gene seems to be widely expressed with ESTs being present for a number of tissues including the lung, liver, kidney, and skin (Unigene entry Mm.30544).

2810451A06Rik has apparent orthologues in all of the higher eukaryotes by reciprocal BLAST, and is syntenic in the mammalian genomes. Its orthologue in *Drosophila melanogaster* is the tracheal system development gene *wurst* (Flybase entry FBGN0030805). The *wurst* gene is expressed in tracheal cells from embryonic day 14 until the end of embryogenesis, as well as in other ectodermal tissues. Mutations in *wurst* result in embryonic lethality, with a similar phenotype to mutations in the *megatrachea* gene [102].

ORF28

ORF28 is a 385 amino acid protein with a J-domain towards its N-terminal. PSORT II predicts that C16orf28 may be located at the mitochondria with a potential signal sequence over the first 25 amino acids of the protein, with a possible cleavage site between residues 25 and 26. It has orthologues down to *Drosophila* by reciprocal BLAST searches. The *Orf28* gene was studied as part of an investigation of mouse orthologues of human chromosome 21 genes [101]. RT-PCR and *in situ* hybridisation analysis suggest that it is has low expression in the embryonic brain, branchial arch and otic vesicle (these results are available on its MGI datasheet accession ID: 2181053).

Wbscr18

Wbscr18 is a 219 amino acid type III J-protein, with its J-domain being towards its N-terminal. Reciprocal BLAST analysis reveals that the Wbscr18 protein is conserved down to *Drosophila*. It appears to be widely expressed with ESTs in multiple organs (Unigene Mm.178012). No additional domains were identified through SMART analysis, however, PSORT II predicts a potential transmembrane domain at the C-terminal of the transcript, as well as a cytoplasmic localisation.

3.2.3 The mouse family of J-proteins

This analysis has provided a starting point for the study of the J-proteins in mouse and allows a number of observations regarding the diversity of the family to be made. The majority of the type I and type II J-proteins are of a similar size with the J-domain at their N-terminal. The type III J-proteins, however, show a massive size difference with the smallest, Dnajc19, being 116 amino acids and the largest, Sacsin, being 3830 amino acids. The position of the J-domain in these type III proteins also varies and can be present anywhere within the protein.

The SMART analysis of each of these J-proteins revealed a number of additional domains, providing an insight into diverse structural features of this family of molecular chaperones. The majority of these additional domains are present within the type III J-

proteins (Table 3), and include the Myb_DNA binding SANT domains in Dnajc1 (Mtj1) and Dnajc2 (Zrf1, Mida1), the cysteine string region of the Dnajc5 group of proteins, the clathrin binding domain of Dnajc6 (auxilin), and the zinc fingers of Dnajc20 and 4930461P20Rik.

Figure 3.3.A shows an alignment of all the J-domains from the J-proteins in mouse as well as that from *E.coli* DnaJ. The positions of the structural features of the J-domain are also included. This analysis supports previous conclusions concerning the conservation of the J-domain in the different sub-types of J-proteins, the type I and type II proteins have greater overall conservation than the type III proteins. This lack of conservation is most pronounced in the loop region of the J-domain, where, in the type III proteins it is generally greater in length and contains a wider variety of of types of amino acid. This is particular true of Dnajc20, which has the longest loop region of the mouse J-proteins, which comprises of 20 residues. The four helices are reasonably well conserved throughout all of the mouse J-proteins, with helix 2 the most conserved. There are individual J-proteins that are more divergent, such as Sacsin, which has a smaller than average predicted J-domain and is divergent in helices 1 and 3, as well as having a gap in helix 2. The functional relevance of these more divergent J-proteins would require further study in order to test their ability to stimulate the ATPase activity of Hsp70 proteins, but might provide useful information regarding the required features of all the J-domains.

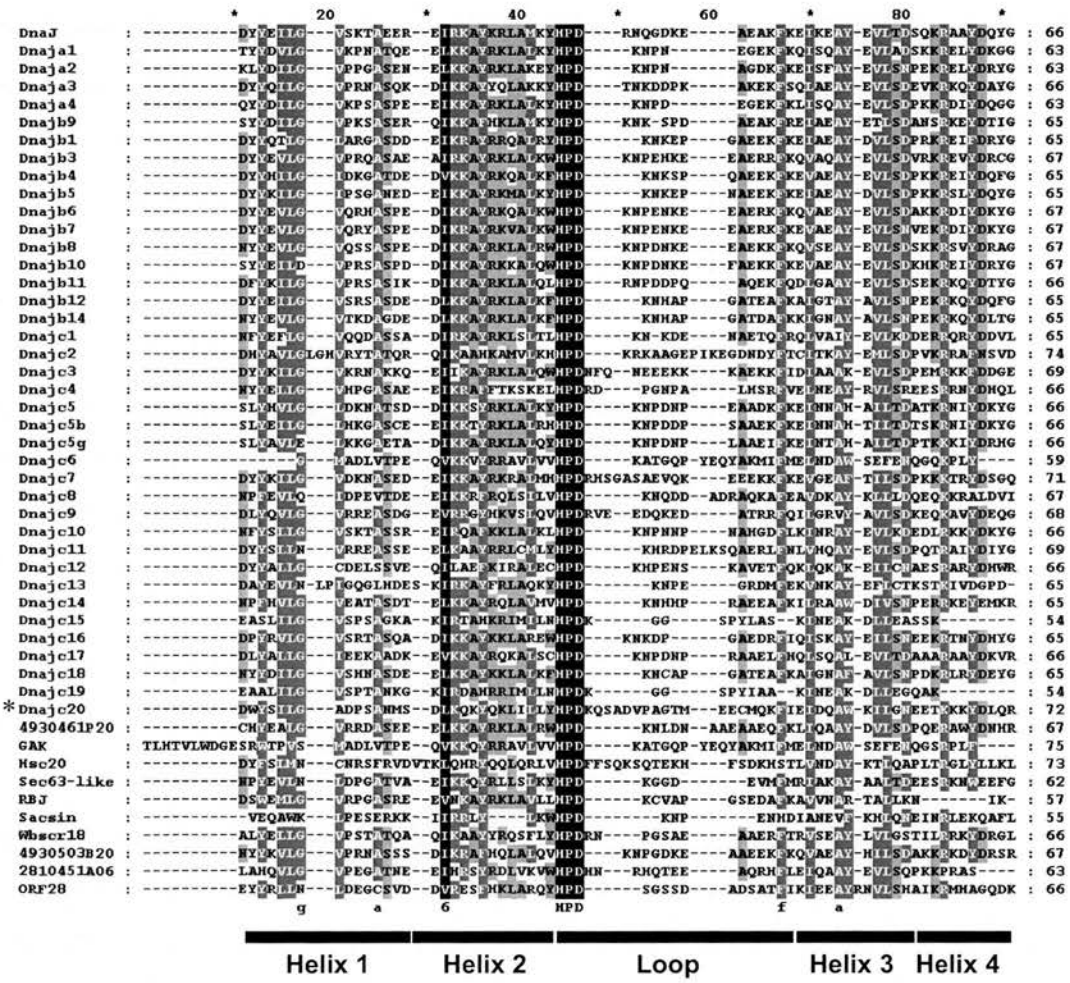


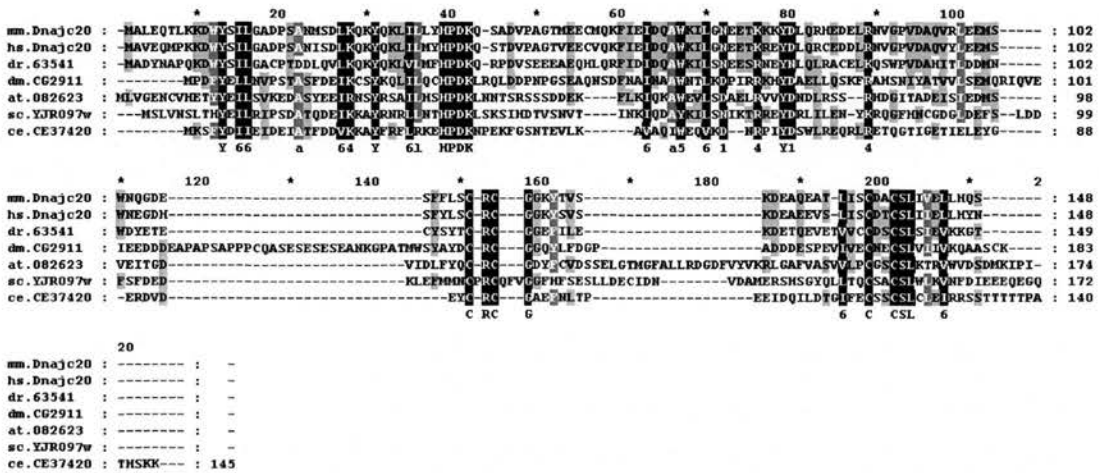
Figure 3.3 Alignment of the mouse J-proteins

An alignment of all of the J-proteins identified in mouse. The structure of the domain is depicted at the base of the alignment. * indicates the Dnajc20 sequence.

3.2.4 The Dnajc20 protein is conserved throughout eukaryotes

Figure 3.4 shows an alignment of the Dnajc20 protein from various eukaryotes. Dnajc20 is most highly conserved in the higher eukaryotes, with 94% similarity between the human and mouse proteins. However, normal BLAST searches fail to identify the sequence orthologues from the lower eukaryotes. The yeast orthologue of Dnajc20, JJJ3 (YJR097w), shares only 27% identity with the mouse protein over their whole length. Using the more sensitive Position-specific iterated and pattern-hit annotated BLAST search facility identified the more distant eukaryotic orthologues by 3 iterations. The conservation in Dnajc20 and its orthologues is highest in the N-terminal J-domain and in a C-terminal domain. Surprisingly, the loop region of the J-domain, which is a defining difference between Dnajc20 and the J-domains of the other mouse J-proteins, shares little conservation across the different eukaryotic Dnajc20 proteins (Figure 3.4.A). There is high conservation in this region between mouse and human, and to some extent down to zebrafish. However, the length of the loop is reasonably conserved between all of the species. The loop region of *Drosophila* CG2911 is the same length as that of the vertebrates, and in the other lower eukaryotes the loop is above average in length. Therefore, the length of the loop may be more important than its amino acid content in a functional context.

A



B

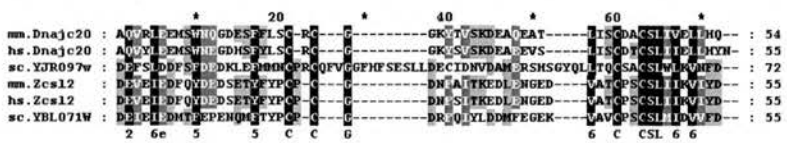


Figure 3.4 Alignments of the Dnajc20 protein in eukaryotes and its homologues

A. An alignment of the Dnajc20 protein in eukaryotes (mm, *Mus musculus*. hs. *Homo* sapiens. dr *Danio rerio*. Dm *Drosophila melanogaster*. At. *Arabidopsis thaliana*. *Sc. Saccharomyces cerevisiae*. Ce *Caenorhabditis elegans*.

B. An aligment of the C-terminal of mouse Dnajc20, its orthologues in humans and yeast, and the homologous proteins from the three species.

3.2.5 The C-terminal domain of Dnajc20

The presence of specific substrate binding domains in some of the type III J-proteins provides important information concerning the functional roles of these co-chaperones. The conservation of the C-terminal region of Dnajc20 across the eukaryotes, particularly the presence of the four cysteine residues (positions 114, 116, 135 and 138 in mouse Dnajc20), suggested that there was a conserved structure in this region, a feature which may be functionally relevant. BLAST search analysis identified an homologous sequence to this C-terminal domain in another protein (Figure 3.4.B). The mouse version of this novel protein is 82 amino acids in length, and is essentially the C-terminal region of mouse Dnajc20. It is conserved throughout eukaryotes, with one copy being present in each species except humans. The human genome appears to contain two such sequences, a three exon gene on chromosome 3, and a one exon gene on chromosome 20. The chromosome 20 gene, however, is annotated as being expressed on the opposite strand to exon 1 of the gene *Tcfl5*. The high conservation between the two human copies suggests there has been a recent duplication, and the chromosome 20 copy is a putative pseudogene.

After the identification of the C-terminal region of Dnajc20, and its homology to another eukaryotic protein, the domain became annotated in the Pfam database (http://www.sanger.ac.uk/Software/Pfam/). It was annotated as a zinc finger CSL

domain, with the proteins that contain it becoming the zf-CSL family. The domain is named because of the motif at the final cysteine. Following this annotation, the proteins in the higher eukaryotes became named, in some databases, as Zcsl1 to 3. The Zcsl3 protein is Dnajc20, Zcsl2 is the homologous protein, and Zcsl1 corresponds to the potential pseudogene in humans. The presence of a defined domain adjacent to the J-domain of Dnajc20 has the potential to offer clues regarding function, in a similar way to the other type III J-proteins with specific substrate binding domains such as Dnajc6 (auxilin). At the time of this annotation only Dnajc20 had been published, and this was concerned with the J-domain. However, over the past two years a number of reports have suggested roles for this domain in a variety of processes, and these will be discussed later.

3.3 Discussion

3.3.1 Dnajc20 and the mouse family of J-proteins

The number of identified eukaryotic J-proteins is rapidly expanding. Recently, Walsh *et al* showed the presence of 22 J-proteins in yeast, making it the first eukaryote to have its full complement of this family of molecular chaperones identified[53]. However, the J-protein family in mouse is complicated by a difficult history of nomenclature, resulting in numerous names for the same proteins, misannotations and omissions. By using a bioinformatic approach I have analysed all of the proteins annotated as containing a J-domain, and have been able to produce a more definitive list of the J-proteins present in the mouse. This analysis has revealed the presence of at least 48 genuine J-proteins, which represents the most conclusive list of J-proteins in a higher eukaryote. Furthermore, this work has revealed the presence of 4 novel mouse J-proteins, and allows the proper characterisation of the Dnajc20 proteins position within the family.

The identification of 48 J-proteins in mouse shows there has been a large expansion of the family in higher eukaryotes. This expansion is restricted to the type II proteins, 12 in mouse compared to 4 in yeast, and the type the type III proteins, 32 compared to 13. In the case of the type I J-proteins, yeast has 5 compared to the 4 in mouse. This difference is probaby due to the fact that two of these type I proteins in yeast, Ydj1 and Xdj1, have high sequence homology and may have similar roles *in vivo*[53]. The large increase in the type II and type III proteins could be due to gene duplications leading to more J-proteins

having analagous functions, or by other proteins co-opting a J-domain and recruiting chaperone power to a novel process. The most likely explanation is a combination of both of these mechanisms. For instance, mouse Dnajc6 (auxilin), the mouse orthologue of yeast Swa2, has high homology over most of its sequence with another J-protein GAK. The GAK protein, which has been shown to have a similar function to Dnajc6[81], is the likely product of a duplication of Dnajc6 followed by the acquisition of a protein kinase domain at its N-terminal. The Dnajc5 (Csp) protein, and its homologues Dnajc5β and Dnajc5γ, do not have have yeast orthologues. These proteins will have evolved after divergence from yeast, and have recruited chaperone power to their role in neurotransmitter release.

The J-domains of these 48 mouse proteins support previous conclusions concerning domain conservation and evolution. The type I and type II proteins show greater conservation in the J-domain than the type III proteins. The reasons for this are generally thought to be the closer evolutionary relationship between the type I and type II proteins, and their need to maintain a functional partnership with their Hsp70 partner[103].

The type III J-proteins overall are most divergent in the loop region of the J-domain, where the length is generally greater and there is a greater variety of amino acid types. This is particularly true in Dnajc20, which has the longest loop region within the mouse J-protein family. The length of the loop region is similar in Dnajc20 and its orthologues,

but shows little conservation from the higher eukaryotes down to yeast. The general explanation for the divergence of the J-domain in the type III J-proteins is their evolution. In the J-proteins that have formed from gene duplication there may have been little pressure to maintain the conservation of the J-domain, due to a likely redundancy with their progenitor[103]. This is probably true for those J-proteins with no functional specificity, however, in the instances where it seems that a J-domain has been co-opted into bringing chaperone power to a novel process, probably via recombination events, the maintenance of a functional J-domain would be beneficial. The *in vitro* ability of Dnajc20 to stimulate the Hsp70 activity of multiple Hsp70 proteins[22], and the *in vivo* ability of the J-domain of Dnajc3 (p58IPK) to substitute for the J-domain of yeast Ydj1 and *E.coli* Dnaj[104], shows that some type III proteins are potentially non-specific in Hsp70 partner choice. A low level of preference for Hsp70 partner in the type III J-proteins might even be expected. If these J-proteins have gained a J-domain by some genetic event, and inherited chaperone power, the Hsp70 partner available might not be the Hsp70 partner to which the J-domain was originally intended. Therefore, there would be room for variation within the J-domain in order to optimise its interaction with a new Hsp70 partner. Recently, Hennessy *et al* suggested a model whereby helices 2 and 3 of the J-domain contact the ATPase domain of Hsp70[105]. This would explain the higher level of conservation in these two regions compared to the loop region, and why mutation in the loop is seemingly tolerated. As the loop region is between helices 2 and 3, mutations here may have a smaller effect the interaction with an Hsp70 protein to lesser extent, and may act to modify the association. This may be an effective

mechanism for the J-domain of the type III J-proteins to optimise there interaction with, an essentially novel, Hsp70 partner.

3.3.2 The C-terminal domain of Dnajc20

The presence of specific substrate binding domains in the type III J-proteins such as Dnajc6 (auxilin), the cysteine string proteins, and Dnajc3 ($p58^{IPK}$), led to the suggestion that other J-proteins may have specialised regions outside of their J-domains that define their function[54]. At the start of this project BLAST searches with the mouse Dnajc20 protein sequence revealed apparent sequence orthologues throughout eukaryotes, which all shared homology in the J-domain as well as in a Cysteine rich region in the C-terminal of the protein. This C-terminal domain of Dnajc20 subsequently became annotated as the zinc finger CSL domain, which defines a family of proteins called the zf-CSLs. The zf-CSL domain is present in two proteins in mouse, Dnajc20 and the 82 amino acid Zcsl2. Both are conserved throughout eukaryotic evolution. At the time of this annotation, however, nothing was known concerning the function of the domain in either Dnajc20 and its orthologues or in the small zf-CSL proteins.

The hypothesis that type III J-proteins have evolved by co-opting a J-domain, and therefore chaperone power, to a novel situation suggests that the function of Dnajc20 will be defined by the zf-CSL domain. Over the past three years a number of reports

have identified a potential role for the zf-CSL domain in multiple functional contexts, including transcription, translation and secretion.

The first evidence regarding the cellular function of the zf-CSL domain was the identification of YBL071w-a, the small zf-CSL protein of yeast, as Kti11, one of a number of factors conferring sensitivity to the *Kluveromyces lactis* toxin zymocin[106]. The *kti* (*K*iller *t*oxin *i*nsensitive) mutations result in the *tot* (*t*arget *o*f *t*oxin) phenotype, which includes resistance to zymocin[107,108]. Several of the *TOT* genes encode members of the Elongator complex[107], including the yeast orthologue of mammalian *Elp4*, the gene adjacent to *Pax6* on mouse chromosome 2[109]. The Elongator complex was initially identified because of its association with hyperphosphorylated RNA polymerase II (RNA pol II)[110]. Elongator possesses histone acetyltransferase (HAT) activity, provided by Elp3, and is thought to facilitate the elongation phase of transcription[111,112]. The human Elongator complex has also been purified and shown to have HAT activity and the *in vitro* ability to interact directly with RNA pol II[113]. Interestingly, the human orthologue of yeast Elp1 was shown to be the IKAP protein. Mutations in IKAP result in the human disease familial dysautonomia, providing a link between Elongator and human disease[114,115].

Biochemical studies have shown that Kti11 is able to directly interact with Elp1, Elp2, and Elp3, leading to the suggestion that KTI11 may mediate the association of Elongator with the transcription machinery[106,116]. Loss of the Kti11 protein results in the

accumulation of a truncated form of Elp1 that cannot enter the nucleus[116]. However, a transcriptional role for the Elongator complex has regularly been questioned. Some reports have failed to identify an association between Elongator and the transcription machinery[117,118], with the subunits having predominantly cytoplasmic localisation[117]. Recently, a novel function for the Elongator complex has been suggested in the regulation of polarised cell-surface transport[119]. The yeast Sec2 protein is a guanine nucleotide exchange factor required for activation of the Rab GTPase Sec4, involved in polarised cell surface trafficking of post-Golgi secretory vesicles. The Elp1 protein was shown to directly interact with Sec2 and is required for its correct localisation. Furthermore, Rahl *et al* gave more conclusive evidence that Elp1 does not enter the nucleus[119]. Previous studies had suggested that the localisation of the Elongator subunits was cytoplasmic, however, they could not rule out nuclear shuttling. Rahl *et al* used a cell line deficient in nuclear export, therefore, any Elp1 shuttling into the nucleus would be unable to return to the cytoplasm, however, no nuclear localisation was observed. The authors note that it is still possible that under certain unknown conditions Elp1 may localise to the nucleus.

The identification that Elongator may function in secretion supports previous findings regarding the human orthologue of Kti11. A yeast two-hybrid screen for proteins that interact with DelGEF (deafness locus putative guanine nucleotide exchange factor) identified the small human zf-CSL protein (Zcsl2) and it was named DelGIP1 (for DelGef interacting protein1)[120]. The DelGEF protein, like Sec2 a guanine nucleotide

exchange factor, is associated with the human orthologue of yeast Sec5, a subunit of the Sec6/8 multiprotein secretion complex[121]. Downregulation of DelGIP1 in HeLa cells resulted in the increased secretion of proteoglycans, thus suggesting a potential role for the zf-CSL domain in secretion[120].

It is unclear whether the identified role for the small human zf-CSL protein DelGIP1 is directly linked to the apparent function of Elongator in secretion. However, both appear to negatively regulate the localisation of guanine nucleotide exchange factors. The findings of Rahl *et al*[119], and the previous work by Krogan[118] and Pokholok[117] contrasts starkly with good evidence that suggests Elongator, and therefore Kti11, are required to facilitate RNA pol II transcription. Further work is required to clarify the biological role of the Elongator complex before its association with the zf-CSL domain can properly be elucidated.

The most clear-cut evidence for a biological role of the zf-CSL domain comes from two recent papers by Liu *et al*[122,23]. Firstly, the use of retroviral insertional mutagenesis identified a role for the small zf-CSL protein in the synthesis of diphthamide, the target of bacterial ADP-ribosylating diphtheria toxin and *Pseudomanas* exotoxin A[24,124]. Secondly, the study of the diphthamide biosynthesis pathway in yeast identified a requirement for both the small zf-CSL protein and the yeast orthologue of Dnajc20[23]. This work is discussed in detail in the next chapter and forms the basis for the investigation of a cellular role for mouse Dnajc20.

Chapter 4

Dnajc20 is required for diphthamide biosynthesis

4.1 Introduction

Diphthamide is a post-translational modification unique to translation elongation factor 2 (eEF2), and is the target of the ADP-ribosylating diphtheria toxin (DT) and *Pseudomonas* exotoxin A (ETA)[24,124]. Although it is 25 years since diphthamide was identified its function in normal cell physiology is still unknown. Mutations in yeast and CHO cells were generated around 20 years ago that fail to make diphthamide, however, it is only in the past year that all of the proteins have been identified[125-127]. Analysis of both the yeast and CHO mutant cells has revealed the requirement of both the small zf-CSL protein (Dph3/Kti11/YBL071w-a in yeast and DESR1 in CHO cells), and the yeast orthologue of mouse Dnajc20 in diphthamide biosynthesis (JJJ3/YJR097w).

4.1.1 DT and ETA inhibit protein synthesis

Diphtheria toxin (DT) and *Pseudomonas* exotoxin A (ETA) belong to the family of ADP-ribosylating toxins that includes cholera, pertussis and the C2 and C3 toxins of *Clostridium botulinum*. These toxins catalyse the transfer of ADP-ribose from NAD+ onto a target protein, and result in the disruption of various metabolic processes within the host cell. Although the action of this family of toxins is the same, the target varies; the *Clostridium* C2 toxin target is G-actin, the C3 toxin target is Rho, and the cholera and *pertussis* toxin targets are the G-proteins. The target of both DT and ETA is the translation protein eEF2, specifically the modified histidine residue diphthamide[128]. The

eEF2 protein is essential for polypeptide elongation in protein synthesis, forming a complex with GTP and the ribosomes and catalysing the GTP dependent translocation of peptidyl-tRNA from the aminoacyl (A) site to the peptidyl (P) site on the ribosome (Figure 4.1). The GTP binding domain is at the N-terminal end of eEF2 and the interaction with the ribosome is towards the C-terminus. The DT and ETA toxins bind to the same region of eEF2 as the ribosomes and catalyse the transfer of ADP-ribose from NAD^+ to the diphthamide residue, inactivating eEF2 and causing a lethal blockage in protein synthesis[24].

4.1.2 Diphthamide biosynthesis

The diphthamide modification is present in all eukaryotic organisms, where it is restricted to a histidine residue of eEF2 (position 715 in mammals and 699 in yeast)[129]. There is also evidence of a diphthamide modification, or a precursor of diphthamide in the analogous elongation factor of archaebacteria[130]. However, it is absent from EF-G, the eubacterial orthologue of eukaryotic eEF2. Therefore, DT and ETA are specific for the eukaryotic, and perhaps archaebacterial, protein synthetic machinery, but do not target the elongation factor of the pathogens that produce them. Even though it is over 25 years since diphthamide was identified little is known about its function.

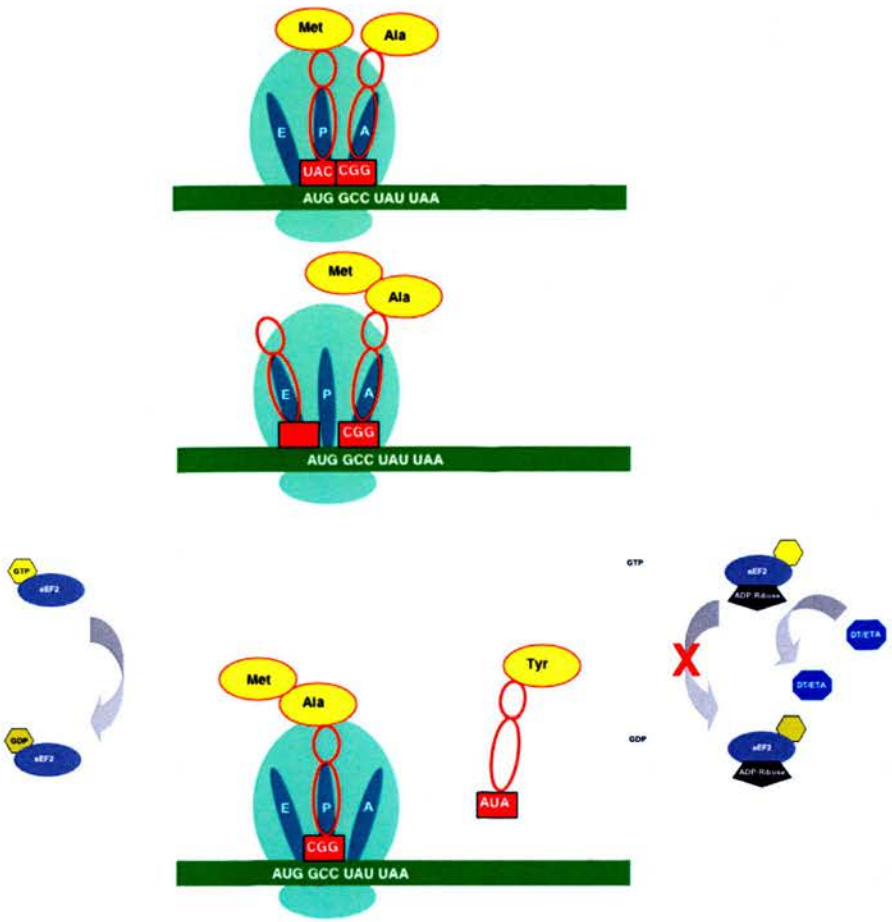


Figure4.1 Translation elongation is inhibited by DT and ETA

(Adapted from http://nobelprize.org/chemistry/educational/dna/a/translation/elongation.html)

eEF2 catalyses the translocation of peptidyll-tRNA from the A site to the P-site in a GTP dependent process in the final step of translational elongation. DT and ETA bind to the region of eEF2 that interacts with the ribosome and catalyse the addition of an ADP-ribose moiety to the diphthamide residue. This removes the ability of eEF2 to interact with the ribosome, inhibiting protein synthesis.

Diphthamide deficient mutations have been generated in yeast[127] and CHO cells[125] by chemical mutagenesis and selected due to their resistance to diphtheria toxin. The mutations fall into 5 complementation groups in yeast termed Dph1, Dph2, Dph3, Dph4, and Dph5[127], and 4 in CHO cells termed CG-1, CG-2, CG-3, and CG-4[23,125]. However, the mutant cells, grow at a normal rate and display no obvious phenotypes besides toxin resistance[126,127].

The identification of mutations defective in diphthamide biosynthesis has allowed genetic analysis of the pathway, and revealed it to be one of the most complicated post-translational modifications identified[122]. Figure 4.2 illustrates the diphthamide biosynthesis pathway, and the steps affected by each of the yeast and CHO complementation groups[23,125,127]. The first step involves the transfer of a 3-amino-3-carboxypropyl moiety from S-adenosylmethionine (AdoMet) to the C-2 imidazole of the target histidine to produce an intermediate structure, and requires Dph1 to 4 and CG-2 to 4. Subsequently, trimethylation of the intermediate modification by an AdoMet-dependent methyltransferase to produces diphthine, a process requiring Dph5/CG-1[131-133]. The final step is the amidation of the side chain carboxyl group of diphthine by an ATP-dependent enzyme resulting in diphthamide. No mutations have been identified in this final step as diphthine can also be ADP-ribosylated, although at a slower rate than diphthamide[126], and an ATP-dependent amidation enzyme is still to be identified[23]. Finally, DT and ETA are able to ADP-ribosylate diphthamide at position C-3 of the imidazole ring via transfer of ADP-ribose from NAD^+.

Figure 4.2 The diphthamide biosynthesis pathway

From Liu *et al*, Molecular and Cellular Biology (2004)[23].

The synthesis of diphthamide and ADP-ribosylation. AdoMet dependent 3-amino-3-carboxypropyl transfer on to histidine requires Dph1 to 4/CG-2 to 4, and produces the intermediate state. Dph5/CG-1 catalyses the transfer of 3 methyl groups from AdoMet on to the intermediate structure to produce diphthine. This is followed by an amidation step to produce diphthamide. ADP-ribosylation of diphthamide by DT or ETA results in ADP-ribosyl diphthamide. The relative charge of each modified adduct are indicated above each modification

The Dph2 protein, consisting of 594 amino acids with no known domains, and AdoMet-dependent methyltransferase Dph5 protein were identified by Mattheakis *et al* over 10 years ago[132,134]. However, it is only in the last two years that the proteins mutated in the remaining complementation groups have been identified. Using retroviral insertional mutagenesis, a DT and ETA resistant CHO cell line was identified. The mutated gene encoded the small zf-CSL protein and was named DESR1, for DT and ETA sensitivity required 1[135]. The mutation was shown to correspond to the original CG-2 CHO cell complementation group. Subsequently, the orthologous gene in yeast, Kti11/YBL071w-a was identified as the Dph3 mutation[23].

Transposon mediated mutagenesis was used in the identification of the remaining yeast Dph mutations[23]. Transposon mutagenesis allows the easy identification of the mutated genes by sequencing the transposon insertion site, compared to the original chemically generated mutations, which require mapping. Following selection of DT resistant strains, complementation testing against the original Dph mutant strains was performed. The Dph1 protein was identified as YIL103w a 425 amino acid protein that shares 49% homology with Dph2[23]. The mammalian orthologue corresponds to the CG-4 complementation group in CHO cells[23]. Dph1 and Dph2 were shown to interact by co-immunoprecipitation. Interestingly there is an archael orthologue of the Dph1/Dph2 proteins, which might represent ancient evolutionary conservation of the pathway[148]. The Dph4 mutation was identified as JJJ3/YJR097w, the yeast orthologue of mammalian Dnajc20[23]. However, there is no corresponding CHO cell mutation.

The mechanism of how the Dph1 to 4 proteins function in the first step of diphthamide biosynthesis is unknown. The presence of multiple proteins in the first step of the diphthamide biosynthesis pathway led to the suggestion that they may function as part of a multiprotein complex[132]. This is supported by high-throughput yeast two-hybrid screens in both yeast and *Drosophila*, which show interactions between Dph1 and Dph2, and the co-immunoprecipitation experiments showing interactions between yeast Dph1 and Dph2, and mouse Dph1 and Dph2. Furthermore, recent work in yeast has identified that Dph3/Kti11/YBL071w-a can bind eEF2, Dph1 and Dph2, as well as ribosomal proteins Rps7a and Rps19a[136]. However, as the yeast J-protein, Dph4/JJJ3/YJR097w, was not detected as a Dph3/Kti11/YBL071w-a interactor in the same assay, it has been suggested that it either acts separately to the other Dph proteins, or indirectly on the first step of diphthamide biosynthesis[23]. However, as Dph3/Kti11/YBL071w-a contains little more than the zf-CSL domain, which is shared with Dph4/JJJ3/YJR097w, it is possible that both proteins share the same interactors, but are required at different times.

The identification of JJJ3/YJR097w as Dph4 and its requirement for diphthamide biosynthesis strongly suggested that Dnajc20, its mouse orthologue, may function in the same process and that the Dnajc20 mutant phenotype is due to a deficiency in diphthamide biosynthesis and protein synthesis.

4.2 Results

4.2.1 Dnajc20 Antibody Production

To characterise the function of Dnajc20 and its role in diphthamide biosynthesis a rabbit polyclonal antibody was raised against bacterially expressed full-length Dnajc20. The coding region of *Dnajc20* was amplified by RT-PCR using primers engineered with *Xba1* and *HindIII* restriction sites, and directionally cloned into the pGEX-KG expression plasmid. The pGEX-KG/Dnajc20 plasmid was transformed into calcium competent BL21 *E.coli* cells for the production of GST (glutathione s-transferase)-Dnajc20 fusion protein. The recombinant GST-Dnajc20 protein was purified by incubation with glutathione beads. The Dnajc20 protein was purified from the beads by incubation with thrombin, and the supernatant containing Dnajc20 was sent to Diagnostics Scotland (Edinburgh, UK) for antibody production. Figure 4.3 shows the detection of multiple bands using the antiserum on protein preparations from a wild-type and a mutant mouse embryo subjected to SDS-PAGE and Western blotting (1 in 5000) and a band corresponding to the predicted 17kDa Dnajc20 protein is readily identifiable in the wild-type sample. However, this band is absent in protein from *Dnajc20* mutant embryos (Figure 4.3). Furthermore, protein from mutant embryos contains no band corresponding to the 15kDa of the predicted mutant protein.

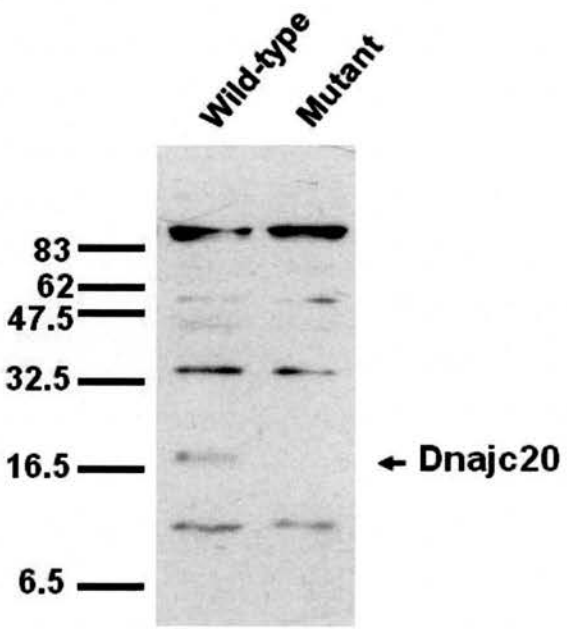


Figure 4.3 Western blot analysis against Dnajc20

Figure 4.3 shows the result of a Western blot using rabbit antiserum produced against the full length Dnajc20 protein following SDS-PAGE of protein samples from wild-type and mutant *Dnajc20* embryos. Multiple bands are detected, including a band of approximately 17kDa that corresponds to the expected size of wild-type Dnajc20 protein (arrow). This band is absent in the mutant protein sample.

The absence a band corresponding to the 15kDa mutant Dnajc20 protein suggests that the mutant transcript is destroyed or, alternatively, the epitope recognised by the antibody may correspond to the region encoded by exon 4 of the *Dnajc20* transcript, and therefore is not present in the mutant protein. Further analysis will be required to determine whether the putative mutant protein is ever synthesised.

The rabbit antiserum is a useful tool for the further investigation of Dnajc20, with the 17kDa wild-type Dnajc20 protein being easily detected by SDS-PAGE and Western blotting. However, the detection of additional proteins means that purification of Dnajc20 antibodies from the antiserum will be required before other possible applications are explored.

4.2.2 Homozygous *Dnajc20* mutant embryos lack diphthamide

To determine whether eEF2 in the *Dnajc20* homozygous mutant embryos had the diphthamide modification, protein was prepared using modified RIPA buffer from each embryo from an E10.5 litter (from an intercross from the CD1 background animals, see Chapter 2). Fresh protein preparations were used as repeated freeze-thaw cycles have been reported to cause the loss of diphthamide from wild-type samples, an event confirmed by our own observations. The protein preparation was subjected to both Native and SDS-PAGE and the eEF2 protein detected by Western blotting (Figure 4.4.A and B). It has previously been shown that eEF2 with the diphthamide modification

carries an additional positive charge compared to unmodified eEF2, causing it to migrate more slowly by Native-PAGE (Figures 4.2 and 4.4.B)[135]. The genotype of each embryo was confirmed by sequencing, using DNA prepared from the embryonic egg sac. The protein from mutant embryos is in lanes 1, 3, 4, 5, and 9, lane 2, 7, and 8 are heterozygous for the mutation in *Dnajc20*, and lane 6 is wild-type. eEF2 protein from the *Dnajc20* homozygous mutant embryos has migrated more quickly either than that of the heterozygous or wild-type embryos, illustrating that the mutant embryos are deficient in diphthamide biosynthesis. SDS-PAGE and Western analysis was also performed against the Dnajc20 protein in the same preparations using the rabbit antisera produced against the Dnajc20 protein (Figure 4.4.C). A band corresponding to the wild-type 17 kDa Dnajc20 protein is present in the protein extracted from both wild-type and heterozygous embryos but is absent in mutant embryos.

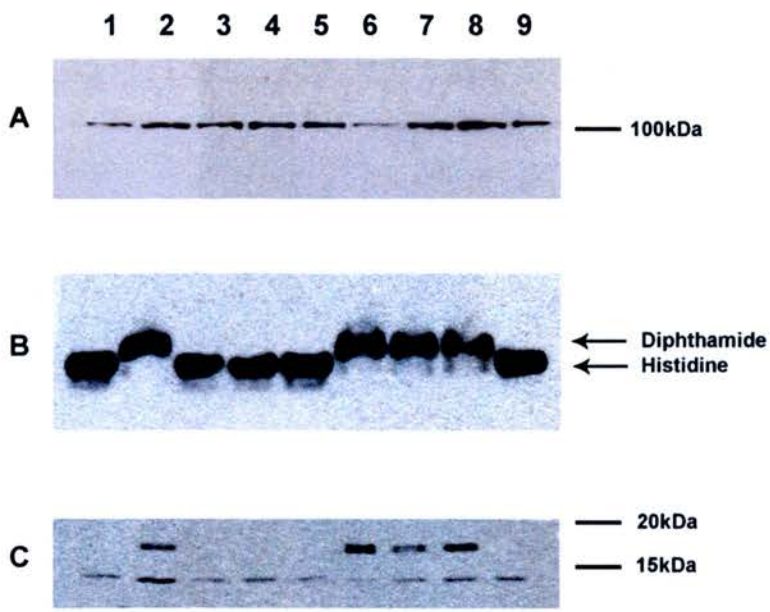


Figure 4.4 Investigation of diphthamide biosynthesis

4.4.A Detection of eEF2 following SDS-PAGE. 4.3.B. Detection of eEF2 following Native-PAGE. 4.4.C. Detection of Dnajc20 following SDS-PAGE. The same samples were run in lanes 1 to 9 in the 3 experiments. Genotype Lane 1 Mutant (Mut), lane 2 Heterozygote (Het), lane 3 Mut, lane 4 Mut, lane 5 Mut, lane 6 wild-type, lane 7 Het, lane 8 Het, and lane 9 Mut. In the mutant samples the eEF2 protein migrates faster indicating diphthamide residue is not present (4.4.B). There is no difference in migration by SDS-PAGE. No Dnajc20 protein (expected size 17kDa) is detected in the mutant samples (4.4.C).

4.2.3 Cellular localisation of the Dnajc20 protein

It has previously been shown that the small zf-CSL protein from yeast Kti11/Dph3/YBL071w-a interacts with diphthamide biosynthesis factors as well as translation proteins. The yeast orthologue Dnajc20, JJJ3/Dph4/YJR097w, was not identified in the complex in this analysis. However, as this protein is required in the same step of diphthamide biosynthesis as Kti11/Dph3/YBL071w-a and they both have the zf-CSL domain; I would predict that they have the ability to interact with the same proteins, but not at the same time. To investigate the position of the mouse Dnajc20 protein in the diphthamide biosynthesis pathway, immunofluorescence experiments were performed using an N-terminal FLAG-tagged Dnajc20 expression construct. To investigate whether Dnajc20 co-localises with the diphthamide biosynthesis machinery the FLAG-Dnajc20 construct was co-transfected into NIH-3T3 with a myc-tagged Ovca1 construct, the mouse orthologue of Dph1 (a gift from Chun-Ming Chen) (Figure 4.5). Unexpectedly, the Dnajc20 and Ovca1 tagged proteins did not co-localise. The Ovca1 protein localised to the perinuclear region of the cell, the same pattern that has previously been published for this construct[137], whereas the Dnajc20 tagged protein displayed a cytoskeletal localisation. Co-staining of transfected cells for the Dnajc20 FLAG construct with phalloidin reveals that the Dnajc20 protein localises to the actin cytoskeleton (Figure 4.6). The different localisation pattern of Dnajc20 and Ovca1 implies that the first step of diphthamide biosynthesis may require events in distinct cellular compartments, and potential roles for Dnajc20 in this process will be discussed

later in this chapter. However, this relies on the localisation of both the tagged Dnajc20 and Ovca1 mirroring the *in vivo* situation. The Dnajc20 antibody detects Dnajc20 by Western analysis, but reveals that it cross-reacts with other protein species. Purification of this antibody may allow the *in vivo* localisation of Dnajc20 to be confirmed.

4.2.4 Initial investigation of Dnajc20 interactions

Co-immunoprecipitation (Co-IP) experiments have been performed in an attempt to identify potential binding Dnajc20 and eEF2. Previous studies have identified a cytoskeletal localisation of eEF2[138-141], providing support for a possible interaction between the two proteins. NIH-3T3 cells were transiently transfected with the FLAG-Dnajc20 fusion protein and binding to endogenous eEF2 investigated by immunoprecipitation (IP) with anti-FLAG antibody and detection of eEF2, or IP with anti-eEF2 antibody and detection of FLAG, however, no interaction was identified. It is possible that Dnajc20 interacts with unmodified eEF2 and is unable to bind eEF2 with the post-translational diphthamide modification. This possible situation can be investigated by either Co-IP using cells deficient in diphthamide biosynthesis or by the production of a mutant eEF2 construct that is unable to be modified.

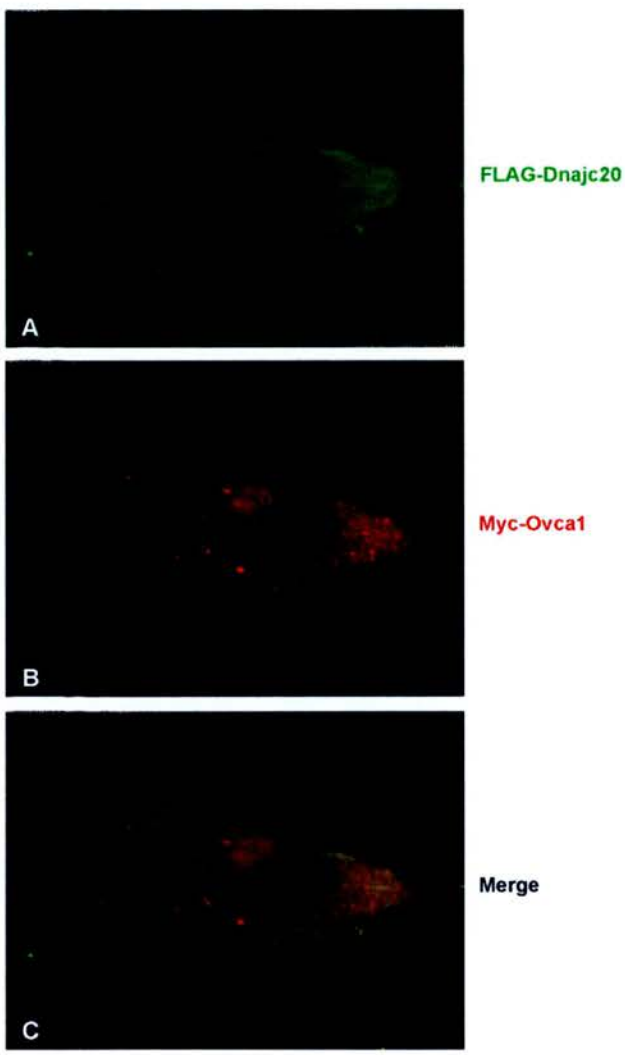


Figure 4.5 Localisation of FLAG-Dnajc20 and Myc-Ovca1

4.5.A to C NIH-3T3 cells transfected with co-transfected with FLAG-Dnajc20 and Myc-Ovca1. 4.4.A staining with anti-FLAG antibody. 4.5.B Staining with anti-Myc antibody. 4.5.C The FLAG-Dnajc20 and Myc-Ovca1 tagged proteins do not co-localise. The FLAG-Dnajc20 protein localises to the cytoskeleton and the Myc-Ovca1 protein localises to a region outside of the nucleus.

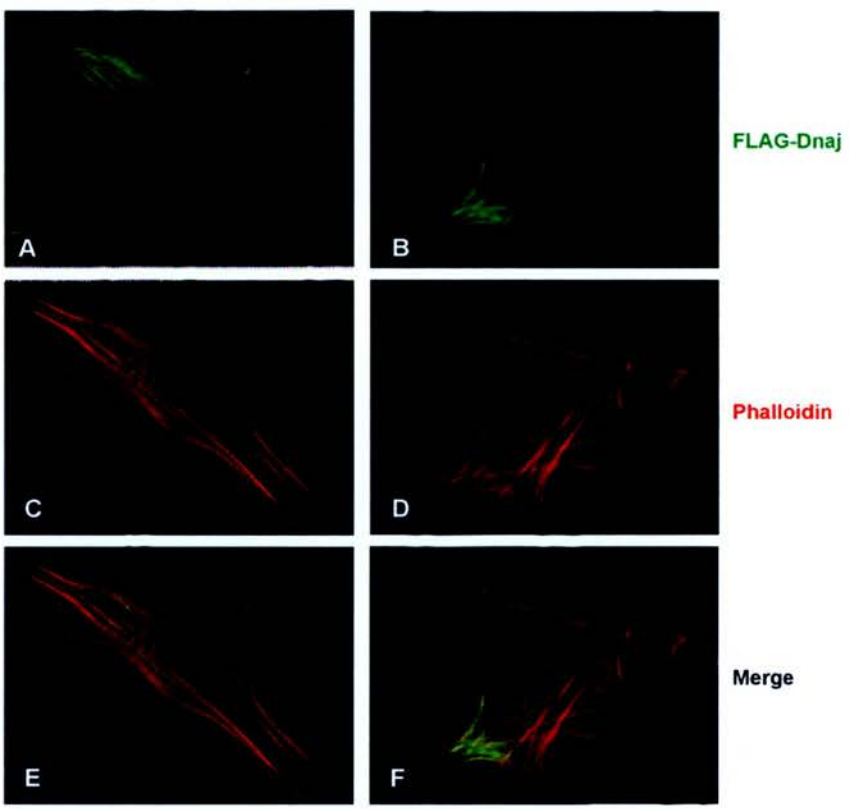


Figure 4.6 FLAG-Dnajc20 localises to the cytoskeleton

4.6.A to F, staining of NIH-3T3 cells transfected with FLAG-Dnajc20, with anti-FLAG and Phalloidin. 4.6.A and B staining with anti-FLAG. 4.6.C and D. staining with Phalloidin. 4.6.E and F merge. FLAG-DNAJ localises to the actin cytoskeleton.

Co-IP experiments were also attempted to identify an interaction between the Dnajc20 and Ovca1 fusion constructs. However, the Myc-Ovca1 fusion protein was never detected by IP. This raises the possibility that Myc-Ovca1 is misfolded and is not a functional fusion protein, a potential explanation for its different localisation to FLAG-Dnajc20.

Further investigation of possible Dnajc20 interactions is still required. Even though no interaction has been observed between Dnajc20 and either eEF2 or Ovca1, the reasons stated above offer possible explanations for the failure to identify binding. Co-IPs with the other members of the diphthamide biosynthesis pathway are also required to help determine the role of Dnajc20 in this pathway.

4.3 Discussion

Diphthamide, the cytosolic target of diphtheria toxin (DT) and *Pseudomonas* exotoxin A (ETA), is a unique post-translational modification restricted to eEF2. The identification that the yeast orthologue of Dnajc20, and both the mammalian and yeast small zf-CSL protein, are required for diphthamide biosynthesis led to the investigation of whether the same is true in mouse. The use of Native-PAGE revealed that eEF2 from *Dnajc20* homozygous mutant mice migrated more quickly than eEF2 from wild-type or heterozygous littermates, due to a charge difference between diphthamide and the unmodified histidine residue. This illustrates that the Dnajc20 protein is required for the first step of diphthamide biosynthesis in mouse.

The first step of diphthamide biosynthesis involves the transfer of a 3-amino-3-carboxypropyl moiety from S-adenosylmethionine (AdoMet) to the C-2 imidazole of the target histidine, and requires 4 proteins in yeast, Dph1 to Dph4. The mechanism by which these proteins catalyse this transfer is unknown. However, Dph3, the small zf-CSL protein, has been shown to interact with Dph1, Dph2 and eEF2, as well as other translation proteins[136]. Interactions between Dph1 and Dph2 (and their orthologues) have been shown in yeast and *Drosophila* in global yeast-two hybrid screens, and by co-immunoprecipitation in yeast and mouse. The yeast Dph4 protein, the orthologue of Dnajc20, has not been identified in these experiments.

The zf-CSL domain is present in both Dph3 and Dph4 and their orthologues in the higher eukaryotes. As the only conserved region present in the small zf-CSL proteins is the zf-CSL domain itself, it is likely that the interactions with the other diphthamide biosynthesis proteins and translation proteins would be mediated by this domain. Therefore, it was hypothesised that the J-domain zf-CSL proteins might be able to interact with the same group of proteins as the small zf-CSL proteins, but at a different time. To investigate this, initial experiments to show the localisation of the mouse Dnajc20 protein were performed. Tagged Dnajc20 protein localised to the actin cytoskeleton, whereas tagged Ovca1, the mouse orthologue of Dph1, displayed a perinuclear localisation. The different localisation pattern of tagged Dnajc20 and Ovca1 implies that the first step of diphthamide biosynthesis may take place in two distinct cellular compartments. If this was the case, an event requiring Ovca1/Dph1, and presumably Dph2 and the small zf-CSL protein Dph3, would occur in the perinuclear region, and a second event involving Dnajc20/Dph4 would take place at the cytoskeleton. Alternatively, the Dnajc20 protein may not be directly involved in the 3-amino-3-carboxypropyl transfer in the first step of diphthamide biosynthesis, and instead, may provide a chaperone function in either the folding of one of the components of the pathway or the formation of the required complexes.

Interestingly, the localisation of Dnajc20 with the actin cytoskeleton may suggest a direct link to the eEF2 protein. A number of the translation elongation factors, including eEF2[138-141], eEF1alpha[142] and eEF1beta[143], have been implicated as actin binding

proteins and cross-communication between the protein synthetic machinery and the cytoskeleton has been suggested. Initial attempts have been made to identify co-localisation of Dnajc20 with eEF2, however, these have been unsuccessful, and further analysis of the cellular compartmentalisation of the eEF2 protein is required. Co-immunoprecipitation experiments have also been performed in an attempt to identify potential binding Dnajc20 and eEF2, but results to date suggest that there is no interaction. The results of Native-PAGE suggest that almost the entire eEF2 protein population has the diphthamide modification, and there is very little of the unmodified precursor (Figure 4.3.B). As Dnajc20 is required for diphthamide biosynthesis, any interaction with eEF2 may be restricted to the unmodified precursor. The same experiment using cells that are deficient in diphthamide biosynthesis would allow the further study of the potential interaction between Dnajc20 and unmodified eEF2, and the function of Dnajc20 in the first step of diphthamide biosynthesis.

Localisation of the tagged Dnajc20 protein to the actin cytoskeleton suggests that the role of Dnajc20 in diphthamide biosynthesis could be a result of association with eEF2. There are a number of mechanisms that Dnajc20 could perform in this putative role, based upon a requirement for a chaperone. It may be required for the folding of eEF2 in to a structure that allows the first step of diphthamide biosynthesis to occur. Alternatively, it might be involved in the correct localisation of eEF2, or other required factors. Another possibility might be in the protection of the intermediate structure produced in the first step of diphthamide biosynthesis. Results obtained in the same

high-throughput yeast two-hybrid analysis that revealed an interaction between Dph1 and Dph2 provides another possibility in identifying the role of Dnajc20. In this screen the yeast orthologue of Dnajc20, Dph4/JJJ3/YJR097w was found to physically interact with an uncharacterised protein called YNL092w[144]. This protein is an outlier of the AdoMet-dependent methyltransferase family, and it is tempting to speculate that this enzyme might be involved in the 3-amino-3-carboxypropyl transfer from AdoMet to the target histidine residue in the first step of the diphthamide biosynthesis pathway. The orthologous protein in mouse, 2410127L17Rik, shares 50% similarity, but is also uncharacterised. However, this interaction may not represent the true physiological situation, and further analysis is needed to firstly confirm any potential interaction, and secondly, identify a requirement for this protein in diphthamide biosynthesis.

The presence of mouse models for diphthamide biosynthesis will provide valuable tools for the study and eventual elucidation of its significance in normal cell physiology. The *Dnajc20* gene is the second diphthamide biosynthesis factor for which a mutation exists. The mammalian orthologue of the *Dph1* gene is *Ovca1*, a ubiquitously expressed gene first identified as a potential tumour suppressor in humans[145]. The total or partial loss of *Ovca1* expression has been observed in ovarian tumour tissues and cell lines[145,146]. Due to this association with ovarian cancer, a knockout *Ovca1* mouse has previously been generated[26]. The homozygous phenotype of the *Ovca1* null mouse is very similar to ENU generated Dnajc20 mutant animals.

Ovca1 null mice on a B6/129 background die soon after birth with a body weight around 50% of their wild-type and heterozygous littermates[26]. Furthermore, a proportion of *Ovca1* null embryos die around midgestation, a situation similar to that with the *Dnajc20* homozygous mutant embryos. The severity of the *Ovca1* null phenotype is also background dependent; on a 129/SvEv genetic background all embryos die by E13.5. *Ovca1* null embryos display developmental delay of up to 24 hours compared to wild-type and heterozygous littermates. Figure 4.7.A and B illustrate the E9.5 and E11.5 *Ovca1* null embryonic phenotype and Figure 4.7.C shows the *Dnajc20* null embryos. The most striking phenotypic similarity between the *Dnajc20* mutant embryos and the *Ovca1* null animals is preaxial polydactyly of the hind limbs (Figure 4.7.D to F).

There are phenotypic differences between the *Dnajc20* and *Ovca1* homozygous mutants. All of the *Ovca1* null mice that were born had cleft palates, an abnormality not observed in the *Dnajc20* mutant animals. However, differences such as this could be due to the genetic backgrounds. There are additional features that have been observed in the *Ovca1* mutants, including focal liver damage in around half of null embryos and an immature lung phenotype, which need to be investigated in the *Dnajc20* homozygotes.

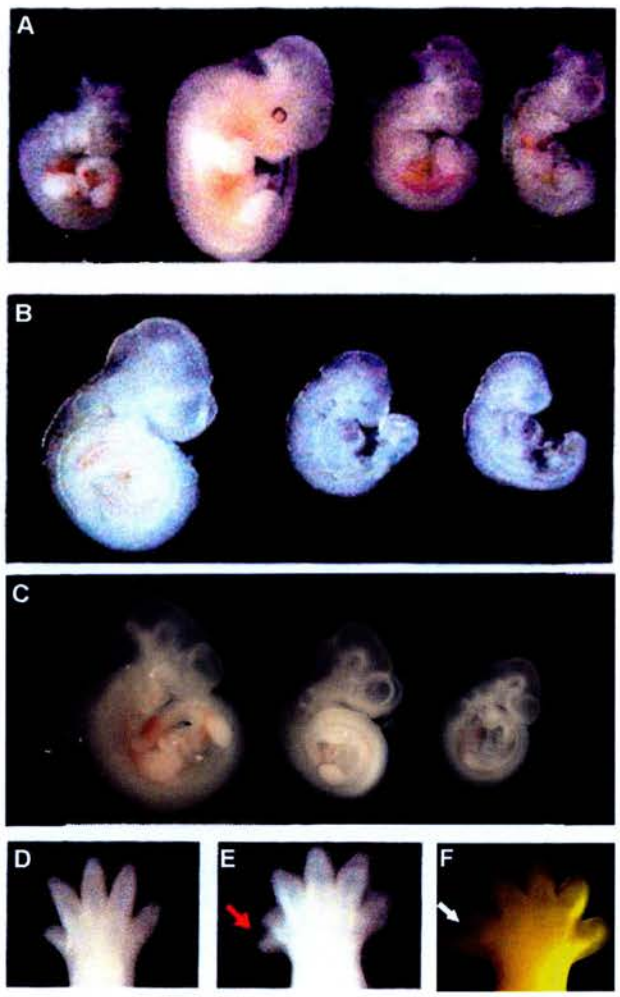


Figure 4.7 *Dnajc20* mutant embryos are similar to *Ovca1* mutant embros

Figure 4.7.A,B,D and E are from Chen and Behringer, Genes & Development. (2004).[26]
4.7.A. E10.5 Wild-type embryo, E11.5 wild-type embryo, and two E11.5 *Ovca1* mutant embryos. 4.7.B. E9.5 Wild-type embryo, and two E11.5 *Ovca1* mutant embryos. 4.7.C. Wild-type E10.5 embryo, and two *Dnajc20* mutant embryos. 4.7.D. Hindlimb of wild-type E14.5 embryo. 4.7.E. Hindlimb of *Ovca1* mutant embryo with preaxial polydactyly (red arrow). 4.7.F. Hindlimb of *Dnajc20* mutant embryo with preaxial polydactyly (white arrow).

The cellular phenotype of *Ovca1* mutant mouse embryonic fibroblasts (MEFs) has been investigated and a defect in proliferation identified[26]. The *Ovca1* null MEFs have a reduced S-phase population compared to *OVCA1* heterozygous MEFs. Analysis of the retinoblastoma (Rb) gene revealed a reduced phosphorylation state in the *Ovca1* null MEFs. Phosphorylation of Rb in late G1 promotes entry into S-phase by releasing the E2F transcription factors, whereas dephosphorylation of Rb results in it binding to E2F transcription factors and repressing their transcriptional activity[147]. The phosphorylation of Rb is consistent with a proliferation defect in the *Ovca1* mutant MEFs[26]. The proliferation defect observed in the *OVCA1* deficient MEFs suggests that OVCA1 is a positive regulator of cell proliferation, or alternatively, the loss of OVCA1 results in the activation of a cell cycle checkpoint. The *OVCA1* gene is present on the short arm of human chromosome 17 close to the cell cycle regulator *p53*, and deletion of both genes is observed in human tumours[26]. To whether the loss of *p53* could rescue the proliferation defect observed in the *OVCA1* mutant MEFs Chen and Behringer generated MEFs deficient for both genes[26]. The close proximity of the two genes required a fortuitous spontaneous recombination event to produce mice with mutations in both *OVCA1* and *p53*. Double mutant MEFs produced from these animals had normal proliferation in comparison to the Ovca1 null MEFs. In embryos with homozygous mutation of both *OVCA1* and *p53* some of the developmental abnormalities seen in the *OVCA1* null embryos, including size and the cleft palate phenotype, were partially improved but the embryonic lethality remained. These results do suggest an interaction between *OVCA1* and *p53* in at least a subset of cell types, probably through the

induction of a p53 cell-cycle checkpoint, however, a p53 controlled checkpoint alone does not completely explain the homozygous *Ovca1* phenotype[26].

Due to the loss of *OVCA1* expression in human ovarian tissue and ovarian cancer cell lines and the loss of both *OVCA1* and *p53* in some tumours, Chen and Behringer investigated tumourigenesis in a variety of *Ovca1/p53* genotypes over a two year period[26]. 50% of the *Ovca1* heterozygous animals tested had developed tumours by two years of age, suggesting that OVCA1 acts as a tumour suppressor. The tumour spectrum included epithelial tumours of the lung and liver as well as lymphoid and stromal derived malignancies, suggesting a global tumour suppressor role[26,148]. However, only one ovarian tumour was identified in the mutant *Ovca1* heterozygotes. The heterozygous loss of *Ovca1* also increased the tumour latency and load of *p53* mutant mice, implying a joint role for *Ovca1* and *p53* in some cancer types[148]. It is unclear whether the tumourigenesis in *Ovca1* heterozygous animals requires the loss of heterozygosity, or if haploinsufficiency of *OVCA1* along with additional factors such as mutation of other tumour suppressors is sufficient for a tumour to develop[148]. The presence of an *Ovca1* conditional mutation will provide further information about its role in tumourigenesis.

It will be interesting to see if *Dnajc20* heterozygotes are predisposed to develop tumours, which would imply that diphthamide itself that acts in tumourigenesis. The *Dnajc20* mutant mice, together, with the *Ovca1* mutant, will provide significant

information in the further investigation of the requirement of diphthamide in translation, mammalian development, checkpoint control and tumourigenesis.

Chapter 5

Analysis of a potential homologous mutation in humans

5.1 Introduction

5.1.1 WAGR Syndrome

The region of mouse chromosome 2 deleted in the *Pax6*$^{Sey\text{-}1H}$ mice is syntenic to human chromosome 11p13. Deletion of either part or all of human chromosome 11p13 results in the phenotypically variable WAGR syndrome (Wilms tumour, aniridia, genitourinary abnormalities and mental retardation), which was identified as a syndrome in the mid 1960s with the recognition that patients with aniridia were of a much higher risk of the childhood kidney cancer Wilm's tumour[149]. Analysis of the WAGR deletion subsequently led to the identification of Wilms tumour gene *WT1*[150], which is also responsible for the genitourinary abnormalities, and *PAX6* which is mutated in aniridia[151]. The gene, or genes, responsible for the mental retardation component of WAGR syndrome are yet to be identified.

Wilms tumour is an embryonal neoplasm of the kidney that occurs at a rate of 1 in 10,000 children usually between the ages of 2 and 4 years, however in children with aniridia the occurrence is 1 in 3[152]. Mutations in the *WT1* gene are responsible for between 10 and 15% of Wilms tumours.

Aniridia is an autosomal dominant disorder of the eye which shows almost complete penetrance, although the severity of the disorder varies in its presentation even among

patients with the same mutation, and is characterised by the complete or partial absence of the iris, and is associated with cataract and glaucoma. Aniridia occurs at a rate of approximately 1in 50,000 to 100,000, with around two thirds of the cases being familial, only a small proportion of cases are associated with WAGR syndrome. The heterozygous loss of PAX6 has been shown to be responsible for the majority of aniridia cases, due primarily to nonsense mutations[151,153].

The cause of the mental retardation phenotype, which is a characteristic of many contiguous gene deletion syndromes, has yet to be determined, although the gene, or genes, responsible are thought to map distally to *WT1* and *PAX6*[154]. The desire to identify the cause of the mental retardation has resulted in an increased effort to isolate and characterise the genes within the WAGR deletion region, with the aim of achieving a better understanding of the genes involved in the development and function of the brain[48,154].

In addition to the defining abnormalities of WAGR syndrome other features have also been reported. A small number of WAGR patients have been described with morbid obesity, leading to the proposal of the WAGRO acronym and the suggestion that there is an obesity gene present in the 11p13 region. Other characteristics associated with WAGR syndrome include a small head circumference, thought to be due to the loss of *PAX6*, as well as a low birth weight and body size.

Recently four WAGR patients have been reported with bilateral hallucal polydactyly[20,21]. These reports, along with Malpeuch *et al's* identification of a WAGR patient described as having "hexadactyly and syndactyly of the first and second toes" suggest the presence of a polydactyly gene associated with WAGR syndrome[155]. There is remarkable similarity between the polydactyly WAGR patients and the polydactyly of the *Dnajc20* mutant mice. The appearance of the polydactyly in the *Dnajc20* and *Ovca1* mutant mice and the relative rarity of this type of preaxial polydactyly (OMIM: 174400) strongly implies that the loss of *DNAJC20* is the cause of polydactyly in WAGR syndrome.

5.1.2 The human *Dnajc20* gene in WAGR syndrome

The *WT1* and *PAX6* genes are 700kb apart on human chromosome 11p13, the human orthologue of *DNAJC20* is present a further 400kb telomeric of *PAX6* (Figure 5.1), and would be expected to be heterozygously lost in a large proportion of WAGR patients. Chromosomal aberrations have been mapped in many WAGR individuals, for instance Crolla *et al* (1997)[156] and Crolla and van Heyningen (2002)[157], reported 12 WAGR patients, with variable manifestations of the WAGR clinical features. Of these 5 have lost *DNAJC20* based on the heterozygous loss of the cosmid A4160, which contains the marker *D11S914* (Figure 5.1). In the remaining WAGR patients the cytogenetics was not reported. The mapping information from WAGR and aniridia patients such as these suggests that loss of one copy of *DNAJC20* in humans does not generally contribute to

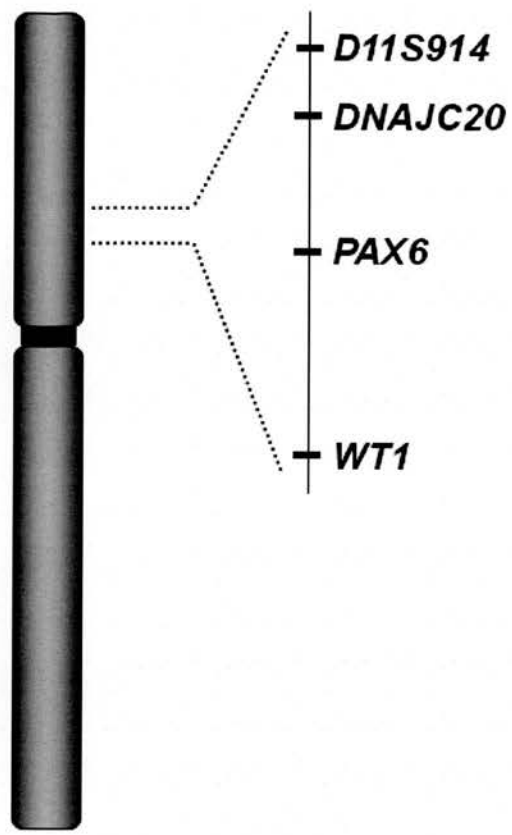


Figure 5.1 Gene order on human chromosome 11

The position of *DNAJC20* in relation to *PAX6* and *WT1* and the *D11S914* marker on human chromosome 11.

the typical manifestation of the syndrome. However, the recent reports of WAGR syndrome patients with polydactyly suggest that there is a gene responsible for polydactyly located contiguously to the WAGR region, and our mouse data makes *DNAJC20* gene the leading candidate.

Individuals have been described with deletions very similar to those in these WAGR with polydactyly patients, as well as deletions that extend further both proximally and distally in WAGR patients with no polydactyly reported. If there is a gene in the WAGR region responsible for hallucal polydactyly the loss of one copy is not generally sufficient to result in the polydactyly phenotype, and like the recessive *Dnajc20* mutation in the mouse, would most likely need to be homozygous. Therefore a second hit in the remaining copy of the gene would be required.

5.2 Results

Manoukian *et al* reported the case of a 30 month old baby girl with bilateral preaxial polydactyly of the feet in addition to the typical clinical features of WAGR syndrome[20]. They identified a deletion, del(11)(p13p14.1), which is between 6.1 and 21.7Mb in length. The human orthologue of *Dnajc20* is lost in the minimal deletion interval.

Bremond-Gignac *et al* reported three patients with hallucal polydactyly of the feet, in addition to WAGR syndrome[21] (Figure 5.1.B). The patients included a pair of

monozygotic twins with bilateral hallucal polydactyly and a third unrelated patient with hallucal polydactyly of the right foot.

Molecular cytogenetic studies of one of the MZ twins (presumed to be representative of both twins) showed a deletion with break points on the boundary of band 11p12 and 11p13 extending to proximal 11p14, with an estimated size between 2.5Mb and 7.0Mb. The absence of the marker *D11S914* shows that human *DNAJC20* is deleted in these patients. Analysis of the third patient revealed a larger deletion del(11)(p13p15.1).

Due to the similarity between the polydactyly in these patients (shown in Figure 5.2) and that observed in the *Dnajc20* homozygous mice it was decided to sequence the second allele of the human *DNAJC20* in order to identify a potential second mutation in the gene. DNA was provided by Radice and Perotti[20] from their patient as well as both of the probands parents. They had previously shown that the deletion carrying chromosome was of paternal origin; therefore a second genetic event within *Dnajc20* would be maternal. Bremond-Gignac[21] provided DNA from the MZ twins, however, they were unable to supply a DNA sample from the third patient.

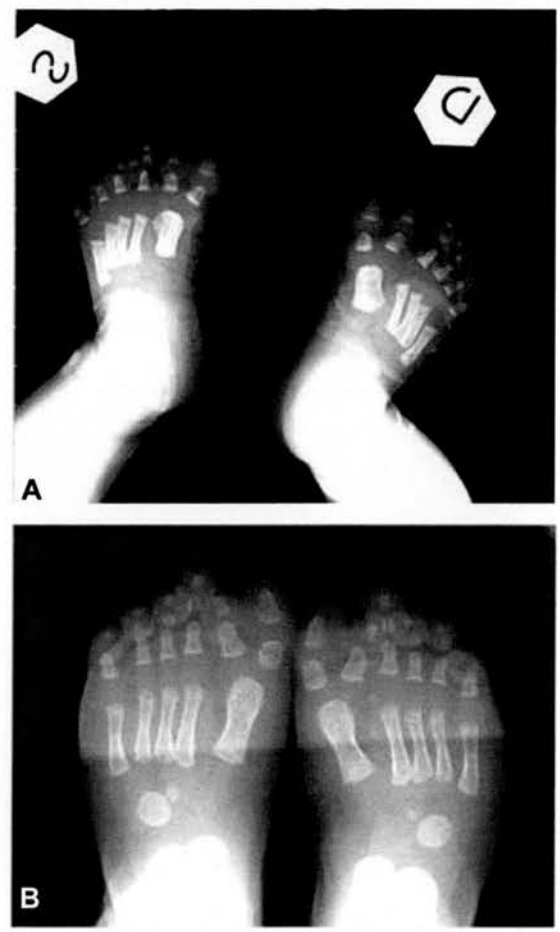


Figure 5.2 WAGR syndrome with polydactyly

Figure 5.2.A. From Manoukian *et al*, American Journal of Medical Genetics (2005).[20]
X-ray showing bilateral hallucal polydactyly of a WAGR syndrome patient.
Figure 5.2.B. From Bremond-Gignac *et al*, American Journal of Medical Genetics (2005).[21]
X-ray showing bilateral hallucal polydactyly of a WAGR syndrome patient, the image is representative of both monoztgotic twins.

Primers were designed to sequence across the exons and splice sites of the human *DNAJC20* gene based on the ensembl annotation. The current ensembl annotation suggests that the gene is encoded by 6 exons, compared to the 5 mouse exons. However the 444bp fifth intron of human *DNAJC20* 444 bp present in the 3 prime UTR contains non-consensus 5 prime and 3 prime splice-sites and was treated as a mis-annotation and was sequenced as being part of the true fifth exon. Additionally, exon 1 which is 456bp in mouse is annotated as being only 26bp in humans, therefore primers were designed so that a homologous sequence to the mouse *Dnajc20* exon 1 was amplified.

PCR and sequencing were performed on the DNA samples of the three WAGR patients with polydactyly, as well as the parents of the proband reported by Manoukian *et al*[20]. No base changes were identified in any of the samples, showing that the hallucal polydactyly is not a result of a second coding or splice mutation in human *DNAJC20*.

5.3 Discussion

The recent reports of WAGR patients with preaxial polydactyly of the feet along with one previous description suggests that polydactyly is a component of WAGR syndrome, although only in a small number of cases. The similarity of the polydactyly in these patients to that in the *Dnajc20* mutant mice is striking, and the human *DNAJC20* gene is an obvious candidate for this additional abnormality of WAGR syndrome. The small number of WAGR syndrome patients with polydactyly, and the loss of the *DNAJC20* gene in WAGR patients without polydactyly, suggested that the manifestation of the abnormality could be due to a second genetic event. However, after sequencing the coding exons, splice sites and 5' and 3' UTRs I have been unable to identify a second allele within the *DNAJC20* gene of these WAGR patients.

The *DNAJC20* gene is still a candidate for the occurrence of hallucal polydactyly with WAGR syndrome. The similarity between the two phenotypes, preaxial polydactyly with duplication of the first digit rather than the more common duplication of the other digits supports this conclusion.

Although no coding change was identified, there is still the potential of a second genetic event. There is a possibility that the polydactyly is caused by a non-coding mutation within the transcript or promoter region of the remaining copy of the *DNAJC20* gene,

however the absence of further tissue from the probands makes the investigation of expression and localisation of the transcript unachievable.

Another explanation, and perhaps the most likely, is that the polydactyly is a low penetrant trait of WAGR syndrome due to haploinsufficiency of *DNAJC20* in some individuals.

The identification of preaxial polydactyly of either one or both feet in 5 individuals with WAGR syndrome suggests that it is a true clinical manifestation of the disorder, albeit only in a small number of cases. The *DNAJC20* gene is the leading candidate for this disorder, with the likely cause either being haploinsufficiency or via a second genetic event although no mutation was identified from sequencing the gene in the three patients.

Chapter 6

Discussion

6.1 Overview

ENU mutagenesis of the mouse provides a method for the identification of novel phenotypes with the advantage that no assumptions are needed about the factors involved, and thus supplies an effective means to identify novel genes and pathways. Region-based screens are a time and cost effective approach in the identification of recessive mutations, especially in the identification of the mutated loci, as the mutation is immediately localised to a specific chromosomal location, sidestepping the need to map the mutation. Our group has recently performed a region-based ENU mutagenesis screen utilising the *Pax6*$^{Sey\text{-}1H}$ deletion, which removes a 3Mb region of mouse chromosome 2, and includes the *Pax6* and *Wt1* genes[47]. This region is of interest for several reasons including; firstly the presence of developmentally relevant factors, which are not explained by known genes within the region, and secondly, relevance to human disease, with deletion of the syntenic region in humans causing the phenotypically variable WAGR syndrome.

At the outset of this project a single mutation had been recovered which is lethal against the *Pax6*$^{Sey\text{-}1H}$ deletion. Intercrosses of carrier animals revealed that the mutation was predominantly lethal when homozygous, however, some homozygotes did survive and these were small with preaxial polydactyly of their hind limbs. The study detailed within this thesis details the further characterisation of the mutant phenotype and, by sequencing exons within the candidate region, the identification of splice mutation in a

functionally novel J-protein. The mutation, a T to A transversion at position +2 of the splice donor site in intron 4, results in the in-frame skipping of the adjacent exon 4 from the mRNA transcript. However, current evidence suggests that either no protein is made, or that the mutant protein is degraded. A bioinformatic approach has been used to clarify the mouse J-protein family in mouse, and led us to suggest the name *Dnajc20* for this gene, in line with the current nomenclature. Bioinformatic characterisation of the Dnajc20 protein identified orthologues throughout eukaryotic evolution, as well as the identification of the C-terminal zf-CSL domain, and its presence in the small zf-CSL proteins, which are also present as one copy in all eukaryotes.

Recent work by Liu *et al* has identified a role for the zf-CSL proteins in the synthesis of diphthamide, the cytosolic target of the bacterial toxins DT and ETA, in yeast and CHO cells[23-25,135]. Consequently, investigation of the *Dnajc20* homozygous mutant mice revealed that they were deficient in diphthamide biosynthesis. *Ovca1*, another diphthamide biosynthesis factor has been mutated in the mouse[26] and shows a very similar homozygous phenotype to the *Dnajc20* mutant mice. The identification of the diphthamide biosynthesis proteins[23], together with the animal models provided by the *Ovca1*[26], and *Dnajc20* mutations provide a major breakthrough in the study of diphthamide, a factor whose function has evaded characterisation since its discovery 25 years ago.

6.2 Diphthamide biosynthesis and its biological relevance

The diphthamide modification was first identified as the target for DT and ETA 25 years ago[25], however, its significance in normal cell physiology is yet to be determined. Mutation of the histidine 699 residue of yeast has implied a structural role for diphthamide either in eEF2 or within the ribosome. Substituting the histidine residue for each of the other 19 amino acids results either non-functional eEF2, illustrating the need for diphthamide in normal translation, or temperature-sensitivity[156], and has led to the suggestion that diphthamide may be responsible for the overall structure and stability of eEF2 [23].

An alternative function of diphthamide is as a site of regulation of protein synthesis. Indeed, there have been previous reports of an endogenous factor that can ADP-ribosylate eEF2, and unlike DT and ETA ADP-ribosylation, would presumably be reversible[161,163]. However, this putative ADP-ribosyltransferase has yet to be identified. There have also been reports suggesting the presence of a cytoplasmic factor that inhibits both DT and endogenous ADP-ribosylation of diphthamide[163,164]. Interestingly, the most recent of these reports propose that the inhibitor is actin[164], which, along with the localisation of Dnajc20 (Chapter 4), provides an additional link between eEF2, diphthamide and the cytoskeleton.

The best evidence for the function of diphthamide comes from recently published work, which has shown that ETA toxin achieves its recognition of eEF2 by mimicry of ribosome structure[159]. Jørgensen *et al* have determined four crystal structures of complexes of yeast eEF2 and the catalytic subunit of ETA toxin[159]. By superimposing this structure onto the previously identified cryo-electron microscope reconstruction of eEF2 in complex with the 80S ribosome[160] they revealed a remarkable similarity between the orientation of ETA and the region of the ribosome that interacts with eEF2[159]. ETA, and presumably DT, by mimicking this ribosomal structure reduce the possibility that a host organism can evolve toxin resistance as coordinated mutations in both the ribosome and eEF2[159]. Interestingly, a possible interaction between diphthamide and two universally conserved adenine residues that are essential for tRNA recognition at the A-site in the small ribosomal subunit was also identified[159]. These two adenines undergo a conformational change, switching from a stacked (closed) position in the absence of tRNA to a flipped out (open) position in the presence of cognate tRNA[159]. Jørgensen *et al* propose that diphthamide may be required for the stabilisation of the stacked position of the adenines, and therefore be essential for maintaining the correct reading frame during translocation across the ribosome[159]. The recent identification of the majority of the proteins required for diphthamide biosynthesis provides a starting point for the further study of diphthamide[23]. An immediate consequence of identifying the diphthamide biosynthesis proteins in yeast and CHO cells has been to assign a biological role for the Dnajc20 and Ovca1 proteins, providing two mouse models for investigation of the physiological relevance of

diphthamide (see chapter 4). The lack of obvious phenotypes in both yeast and CHO cells deficient in the synthesis of diphthamide has made the functional relevance of diphthamide difficult to study[125-127]. Already, with the animal models, it can be concluded that the presence of diphthamide is required for normal physiological development and postnatal survival. The spontaneous generation of tumours in the *Ovca1* heterozygous mice also raises the possibility that diphthamide is linked to cell-cycle checkpoint control and tumourigenesis. If diphthamide is required for the maintenance of the correct reading frame during translation, it might be expected that mutations in the diphthamide biosynthesis pathway result in the synthesis and possible accumulation of mutant proteins, a situation that would be expected to induce checkpoint control.

The mutations in *Dnajc20* and *Ovca1*, as stated previously, supply animal models for the further study of the functional relevance of diphthamide in normal cell physiology. However, how the Dnajc20 (Dph4), Ovca1 (Dph1), Dph2l2 (Dph2), and Zcsl2 (Dph3) proteins function in the first step of diphthamide biosynthesis is still unknown, and a complete knowledge of their role in this synthetic process is required to properly understand the diphthamide modification. Binding studies have previously shown that yeast Dph3 is able to associate with Dph1 and Dph2[136]. It has been suggested that these three proteins work as a multiprotein complex to catalyse the transfer of 3-amino-3-carboxypropyl on to the target histidine of eEF2[23]. The presence of a homologue of the Dph1 and Dph2 proteins in archaebacteria supports this hypothesis. If diphthamide

biosynthesis occurs in the same way in both eukaryotes and archaebacteria it is likely that the Dph1, Dph2 and their archaebacterial homologue provide the catalytic domain required for the first step of the modification. The presence of the zf-CSL domain in Dnajc20/Dph4 and Zcsl2/Dph3 implies that these proteins share a common substrate. The localisation of Dnajc20 (described in chapter 4) and the proteins that have been identified as interacting with yeast Dph3[136] suggest eEF2 as a likely candidate. As described previously, the use of diphthamide deficient mouse embryonic fibroblasts provides a suitable platform for the study of the putative association between eEF2 and Dnajc20.

Another issue concerning the role of Dnajc20 in the synthesis of diphthamide is its function as an Hsp70 co-chaperone. A starting point for the identification of the requirement of chaperone function in the first step of diphthamide biosynthesis will be the identification of the *in vivo* Hsp70 chaperone partner of Dnajc20. The Dnajc20 protein has previously been shown to activate the ATPase activity of mouse Hsp70 proteins BiP and Hsc70[22]. Of these, the constitutively expressed Hsc70 is a good candidate for an *in vivo* partner of Dnajc20 as it has previously been implicated in the organisation of the actin-cytoskeleton[165]. Therefore, to properly characterise the role of Dnajc20 in the first step of diphthamide biosynthesis it will be necessary to identify both its substrate(s) and Hsp70 chaperone partner.

6.3 Dnajc20; a model for type III J-protein evolution

The only region of conservation between the type III subclass of J-proteins and the other family members of molecular chaperones is the J-domain (see chapter 2). Many of these type III J-proteins, along with the J-domain, have specific substrate binding domains that define their function. The structural arrangement and specific functions of these type III J-proteins brought about the hypothesis that J-domains have been co-opted by some proteins through evolution in order to recruit chaperone power to a novel function[54].

The presence of both an orthologue of Dnajc20 and the small zf-CSL protein in the same biological process throughout eukaryotes promote Dnajc20 as a model for the hypothesis of type III J-protein evolution. Structurally it has an N-terminal J-domain and a C-terminal zf-CSL domain. The theory of J-protein evolution suggests an ancient zf-CSL protein will have inherited a J-domain by some genetic event. This new J-domain zf-CSL protein will have recruited chaperone power to a specific process. The identification of a role for both the small zf-CSL protein and the Dnaj20 protein in diphthamide biosynthesis supports this premise. A probable starting point would have been a single gene encoding an ancestral small zf-CSL protein. This gene, via a process of duplication and recombination will have come in to the vicinity of a locus encoding a J-domain, resulting in the presence of both a small zf-CSL and J-domain zf-CSL encoding genes early in eukaryotic evolution. The recruitment of a J-domain to diphthamide biosynthesis, and therefore an Hsp70 protein and chaperone power, will

presumably have improved functional efficiency in the first step of the modification. The maintenance of both proteins in all eukaryotes examined suggests the early evolution and preservation of essential roles for both proteins in diphthamide biosynthesis.

6.4 Is Dnajc20 required for other cellular processes?

The identified role for Dnajc20 in diphthamide biosynthesis is, as discussed, defined by the presence of the zf-CSL domain. However, as described in Chapter 3, this domain is also involved in other biological processes via the small zf-CSL proteins (Chapter 3). This raises the question as to whether Dnajc20 is restricted to the diphthamide biosynthesis pathway, or if it shares more functions with the small zf-CSL proteins in transcription and/or exocytosis via an association with the Elongator complex.

The yeast small zf-CSL protein was originally identified as Kti11, one of the factors resistant to *Kluveromyces lactis* zymocin[106]. The subunits of the Elongator complex have been isolated as the TOT (target of toxin) mutants, which are phenotypically the same as the *Kti11* mutation. The yeast *Dph1* and *Dph2* mutants have also been shown to be moderately resistant to zymocin, suggesting there may be cross-talk between the diphthamide biosynthesis and zymocin resistance pathways[136]. However, mutations in the Elongator complex subunits are not resistant to DT[136]. Therefore, a link between diphthamide biosynthesis and zymocin resistance is not bi-directional. Furthermore, the

zymocin resistance phenotype of *Dph1* and *Dph2* mutant yeast strains could be due to interference with the function of the wild-type Kti11/Dph3 protein. Yeast lacking Dph4/JJJ3/YJR097w have not been tested for zymocin resistance, an experiment that could exclude the possibility of a link to Elongator function.

The Elongator complex is conserved from yeast to humans, and some of the human subunits can substitute for their yeast counterparts, suggesting functional equivalence. However, the role of Elongator complex in the cell is ambiguous, a characteristic that may prove problematic in the investigation of an association between Dnajc20 and Elongator. As discussed earlier, it was originally assigned a role in RNA polymerase II transcriptional elongation via its histone acetylation capacity[110,112,166]. However, whether this is the case has been called into question, for example, there are multiple reports suggesting that Elongator does not enter the nucleus. The alternative role for Elongator in polarised cell-surface transport has been identified more recently[119], and is related to the identification of the human small zf-CSL protein as DelGIP1, downregulation of which, leads to increased proteoglycan secretion[120].

The production of mouse embryonic fibroblasts, homozygous for the *Dnajc20* mutation, will allow initial experiments in the investigation of any potential role associated with Elongator function. For instance, the yeast Kti11 protein has been shown to protect Elp1 from cleavage[136,167], a process that can be easily examined in *Dnajc20* mutant cells. The identification of an association between Dnajc20 and Elongator is therefore achievable,

however, the functional relevance will not be known until the role of Elongator has been elucidated.

6.5 Human DNAJC20

One of the aims of ENU mutagenesis of the mouse is the production of animal models for the study of human disease pathogenesis. Indeed, as stated previously, the ENU mutagenesis screen recently performed by our group against the region of mouse chromosome 2 around the *Pax6* gene has relevance to human disease in the form of WAGR syndrome. The human *DNAJC20* gene is within 400kb of *Pax6*, and one copy is regularly lost in WAGR patients. As discussed in Chapter 5, the similarity between the preaxial polydactyly of the mutant mouse and that recently reported in four WAGR patients makes *DNAJC20* the leading candidate for this abnormality[20,21]. However, sequencing of the remaining copy of the human *DNAJC20* mutations did not reveal any additional mutations. Therefore, if human *DNAJC20* is the cause of polydactyly in these patients it is most likely as a low penetrant trait due to haploinsufficiency. If *DNAJC20* is pathogenic due to hemizygosity, it raises the possibility that it is responsible for other abnormalities in WAGR patients. Another similarity between WAGR patients and the *Dnajc20* homozygous mutant mice is the growth abnormality. Many WAGR patients have a low birth weight and body size. The loss of one copy of *DNAJC20* is a possible contributor to the small size of WAGR patients, however, it is probably a result of the loss of multiple genes within the region.

The potentially most relevant consequence of mutations in human *DNAJC20* is not yet known. As mentioned previously, the phenotype of the *Ovca1* mice links diphthamide to tumourigenesis, and further study of the *Dnajc20* mutant mouse is required for a full understanding putative link between diphthamide and cancer.

6.6 Concluding remarks

It has been 25 years since the cytosolic target of the ADP-ribosylating diphtheria toxin and *Pseudomonas* exotoxin A was identified as diphthamide, a post-translational modification unique to eEF2. However, the function of diphthamide in normal cellular physiology has so far evaded characterisation. Over the past two years there has been a significant breakthrough in the study of diphthamide biosynthesis; first with the identification of the proteins required in its synthesis, and second, the *Ovca1* mutant mouse. The work described in this thesis further contributes to this field of study, showing for the first time the requirement for Dnajc20/Dph4 in diphthamide biosynthesis in a higher eukaryote, and confirming the requirement of diphthamide for normal physiological development and postnatal survival. The *Dnajc20* mutant mouse supplies an additional model for the investigation of diphthamide and its relevance to development, translation, and potentially cell-cycle regulation and tumourigenesis.

Chapter 7

Materials and Methods

7.1 Nucleic acid manipulation

All reagents were analytical grade and supplied by Sigma, Invitrogen, BDH, Roche, Qiagen and Fisher. All nucleic acid manipulation was performed in 1.5ml centrifuge tubes. General lab solutions are prepared by HGU technical service staff, autoclaved, and stored at room temperature unless otherwise stated.

7.1.1 Solutions

EDTA

EDTA (ethyldiaminetetra-acetic acid di-sodium salt) was dissolved in dH_2O, and adjusted to pH 8.0 by adding 5M NaOH.

SDS

SDS (Sodium dodecyl sulfate) crystals are dissolved in dH_2O at 68°C. The solution is adjusted to pH 7.2 by addition of concentrated HCl.

NaCl

NaCl (Sodium Chloride) is diluted in dH_2O.

Tris.HCl

Tris base (tris[hydroxymethyl]aminomethane) was dissolved in dH_2O, and the desired pH achieved by addition of concentrated HCl.

TE buffer

10mM Tris.HCl (pH 7.5), 1mM EDTA.

Quick lysis buffer, 1X

100mM Tris.HCl (pH 8.5)	50ml
5mM EDTA	5ml
0.2% SDS,	5ml
200mM NaCl	20mls
dH_2O	420ml

TBE buffer, 20X

Tris base	216g
Boric acid	110g
EDTA	18.6g

dH_2O added to a final volume of 1 litre. 20X stock solution was diluted to 1X working solution with ddH_2O.

Gel loading buffer, 10X

Reagents are diluted in dH_2O to give a final concentration of 40% Sucrose, 100mM EDTA, and 1% OrangeG.

PCR Primers

Primers were purchased as lyophilised, desalted compounds (MWG Biotech and Invitrogen). Stocks were made up to 1mg/ml with ddH_2O and stored at -20°C. Working stocks were diluted to 100ng/µl with dH_2O and stored at -20°C.

dNTPs

dNTPs were purchased as individual 100mM stocks (Abgene). Working stocks of 10mM were made by mixing 10µl of each dNTP (dATP, dCTP, dGTP, dTTP) with 60µl of dH_2O, and stored at -20°C.

7.1.2 Preparation of DNA

Tissue samples, ear clips or egg sacs, are suspended in 0.5ml Quick Lysis buffer and 2.5µl of 25mg/ml Proteinase K, mixed and incubated overnight at 50°C. The samples are vortexed for 15 seconds followed by centrifugation at 13,000 rpm for 5 minutes. The supernatant is transferred to a fresh 1.5 ml eppendorf, and 0.5ml of isopropanol is added. The samples are mixed and DNA precipitate for 10 minutes at room temperature. The samples are centrifuged at 13,000 rpm for 10 minutes. The supernatant is discarded and

pellet washed with 70% ethanol. The supernatant is removed and the pellet allowed to dry at room temperature. The pellet is then resuspended in 50µl TE buffer or dH_2O.

7.1.3 Preparation of RNA

Total RNA was isolated from embryos using the RNAgents® Total RNA Isolation System (Promega) following the manufacturers instructions. Tissues are homogenised in denaturing solution (Promega). 2M Sodium acetate (pH 4.0) is added to the tissue homegenate, mixed thoroughly, and incubated on ice for 20 minutes. Following incubation, an equal volume of Phenol:Chloroform:Isoamyl alcohol is added to the samples. The samples are then incubated on ice for 20 minutes. The samples are mixed by inversion and centrifuged at 13,000 rpm for 20 minutes at 4°C. The aqueous phase , which contains the RNA, is transferred to a fresh eppendorf. An equal volume of isopropanol is added and samples stored at -20°C overnight. The samples are centrifuged for 30 minutes at 4°C. The supernatant is then removed and the pellet washed with 80% ethanol. The RNA is then dried and re-suspended in nuclease free dH_2O. RNA samples are stored at -70°C.

7.1.4 Determination of DNA and RNA concentration

In order to accurately calculate the concentration of DNA or RNA present in a sample, absorbance in a spectrophotometer at 260nm (A_{260}) was determined. The

spectrophotometer was calibrated using a blank sample containing H_20. Samples were diluted in DEPC treated water, and the absorbance at 260nm was measured. These measures correspond to the following concentrations; 1 A_{260} unit of dsDNA = 50µg/ml H_20 and 1 A_{260} unit of ssRNA = 40µg/ml H_20. Concentration (µg/ml)= A_{260} value x dilution factor x 50µg/ml (DNA)/ 40µg/ml (RNA).

7.1.5 Polymerase Chain Reaction (PCR)

PCR reactions using BIO-X-ACT™ polymerase, 10x OptiBuffer and $MgCl_2$ (Bioline) were carried out to amplify specific regions of genomic DNA in the search for the MUTS1/14 mutation and the sequencing of human Dnajc20 (See Table 7.1). After defrosting on ice, the following reagents were mixed for a single 20µl reaction:

dH_20	14.55µl
10X OptiBuffer	2µl
$MgCl_2$ (25mM)	0.6µl
Forward primer (100ng/µl)	0.6µl
Reverse primer (100ng/µl)	0.6µl
dNTP MIX (10mM)	0.4µl
DNA (50-150ng)	1µl
BIO-X-ACT	0.25µl

Table 7.1 PCR primers used to sequence genes

Exon	Forward Primer 5' to 3'	Reverse Primer 5' to 3'
*Dnajc20*exon 1	TCCTCATCCTTATCACCAACTG	CTCCTGTTTGATGAAGGAGTGG
Dnajc20 exon 2	CTTAAATTTCTGGTCGCACTGA	TACATGGCATTTATGGGACATT
Dnajc20 exon 3	AAACAGTGGCAATAACTGGATG	TCACCACCCTTACTTTTGATTG
Dnajc20 exon 4	GACCCAGAAAGAGGAAAACATT	ATGGCTGCACAAGAGATACATT
Dnajc20 exon 5	TGACGCTGTGTGGAGCCAGGGG	AGCTCTGATGGAGGAGCTCCAC
Wt1 exon 1	TTACTTACCTAGACGGCCTTGC	CTCCTGTTTGATGAAGGAGTGG
Wt1 exon 2	ACGACTCTACTGCTGACGAGAG	AAATGTTTCTACTCGCTCTCCC
Wt1 exon 3	GTCTGATGCTCAGCAAGCTCT	GGATTTCCAGGATCTAGTCTCG
Wt1 exon 4	TCATTATTTTAGAAGGCATGTGCT	TTTTCTAAGCCTAGGCCTGTTC
Wt1 exon 5	TCTTGACGTGGTCATTAGCTTT	GAAGAAGGCAGAGGTAGTACG
Wt1 exon 6	AGTTTTGGATTCTGGCTCCTAT	TTGAATACCTCTTGGTTTGCTG
Wt1 exon 7	AGGGTATGGAAGCAAACTTTTT	ACAACCTTTCTACCCCACAAGT
Wt1 exon 8	AGTACTACCCTAACAGGCTCCC	AAAGAAAAGTCTCCAGGTGAG
Wt1 exon 9	AATCAAAGAAATGTGAGGTCCA	AACGTGAAGAAAAGTTTACGC
Wt1 exon 10	TGGTGATTTAACTCAGGCCTTT	CCCAGTGCTAGTGTTCAGAGTG
1500034J20Rik exon 1	CACAGACCGGGAGCTGAG	CCCAAGAGGTGAGAGAGGC
1500034J20Rik exon 2	TCCTTCCTTAAATTCTCTTTTGG	TGACTATAACAATGTCCCCGC
1500034J20Rik exon 3	*TGGTTATCTTAAGTCATGAACTCAGTC*	TCTGGACTACAGGTGTGTGGAC
1500034J20Rik exon 4	GGCAAAAATTAGAAGCAAAGATAC	AAGACACAGTGACCACAGACAC
1500034J20Rik exon 5	TTTATGACATTGATGTGTTCTGACTC	GCTCAAACTTGGGTCCTCTG
1500034J20Rik exon 6	TTCATATAGGGTTTGTGTAGCTCATC	GGGGAGAGGATAATGTTTTGTC
1500034J20Rik exon 7	ATTGATACACTAAACTGAATTTCTGG	TTTGTTTTGGGTCAATTTTTATC
Elp4 exon 1	GACTTTGGGGCTAGCTCTGAC	AAGAGAGGGAAAAAGGAAATGG
Elp4 exon 2	CACTGAGCACAGTAGTGCAAAA	TGGAGCACAGTGAAGAGTAAAC
Elp4 exon 3	TCTCTTAGGCTCTCTGCTTTCG	AAAACCCAAAAACCCAAAAATC
Elp4 exon 4	TGCTGTCATTA TGGTTCAAGAAA	GGTGGAGAGAGATGGAAGACAC
Elp4 exon 5	CATGGATATGCCATTACTTTGC	TCTTGGTATCCAGGCCTATCAT
Elp4 exon 6	GAAACTAAGTAACTGACCCTCT	TGTGTGCATCCTGAGA ATCTAG
Elp4 exon 7	TCGTCTTCCTGATATTGTGAGT	5' CCAACAAAA ATGGACAAAAACC
Elp4 exon 8	CCTCCACTAAGATTTGTAAGAT	TAAGCCACTGGCACAACAATAC
Elp4 exon 9	TTTAAAGGGAGGATGCT TGTGT	TCATTTGAGTTTGGAGCTTATT
Elp4 exon 10	CTCTCACC ATGTACCCCAGC	AAACTTAAGGATGGTGGCTCC
Hs.Dnajc20 exon 1	AGGAGGAGAGAAGCGACAG	GGGAATCTTCTCCATTCACAG
Hs.Dnajc20 exon 2	ATCTGTGCTTCCTCAGGTTTG	ATTAATACCTGGATTCCTGCC
Hs.Dnajc20 exon 3	TGTTGTCTCATTGCCTTGG	AGGCAACAATAACACTTGGTG
Hs.Dnajc20 exon 4	TGATAGGTAAGTTAGAATTGTAGTGGC	CTCCAGACAGAAACACTTACCAC
Hs.Dnajc20 exon 5.1	TCTGCACCTTTTAAGGGTATTAG	GAAACGTTATCGGTTAGACCAG
Hs.Dnajc20 exon 5.2	TCCCATCTGAGAACTTTATTATGTC	ACAAAACTTAGGCCAGGCAC
Hs.Dnajc20 exon 5.3	TGATATAATATTGGGGCCAGG	TAGTCTACATGCTCAGCCCAC
Hs.Dnajc20 exon 5.4	ACTGCTGCTATGCTGAAATG	CCTAATGGCCACAGATTTTC

After mixing and briefly centrifuging the reactions are placed in a Peltier Thermal Cycler (MJ Research). The standard cycling programme used was:

Step 1 5 minute denaturation, followed by 35 cycles of steps 2 to 4.

Step 2 1 minute at 94°C

Step 3 1 minute at 58°C

Step 4 1 minute 30 seconds at 68°C

Step 5 10 minutes at 68°C, final elongation

Step 6 4°C

This cycling programme works for the majority of primer pairs, however, if PCR failed, the annealing temperature was changed to 55°C and 60°C and repeated.

7.1.6 Reverse-transcription polymerase chain reaction (RT-PCR)

RT-PCR was performed using the Access RT-PCR System (Promega). After defrosting on ice, the following were mixed to a final volume of 50µl:

AMV/*Tfl* buffer (1X)	10µl
dNTPMix (10mM)	1µl
Forward Primer (100µM)	0.5µl
Reverse Primer (100µM)	0.5µl
25mM $MgSO_4$ (100mM)	2µl
AMV Reverse transcriptase	1µl
Tfl DNA polymerase	1µl
Nuclease free dH_20	34µl

After mixing and briefly centrifuging the reactions are placed in a Peltier Thermal Cycler (MJ Research). The standard cycling programme used was:

Step 1 48°C for 45 minutes

Step 2 94°C for 2 minutes, followed by 40 cycles of steps 3 to 5

Step 3 94°C for 30 seconds

Step 4 60°C for 1 minute

Step 5 68°C for 2 minutes

Step 6 68°C for 7 minutes

Step 8 4°C

Table 7.2 shows the primers used for RT-PCR and cloning.

Table 7.2 Primers for RT-PCR and cloning

Primers	Forward primer 5' to 3'	Reverse Primer 5' to 3'
Dnajc20 Exon 3 to 5 RT-PCR	CAAAGTGCAGATGTGCCA	GGAGACAGTGTATTTCCCACC
Dnajc20 (pGEX-KG)	GAAGGCTCCCTTCTAGATCAAAGGACGCTG	TCCTTGTTGTAAAGGCTCGAGGGCAGAAGA
Dnajc20 (p3xFLAG-CMV™-10)	AGGCAAGCTTCTAGGTCAAAGGACGCTGAT	CAGCTGAATTCTGATGGAGGAGCTCCACAA

7.1.7 Gel Electrophoresis

Agarose gels were used to determine the presence and size of nucleic acid molecules. The required amount of High pure agarose (BioGene) was dissolved in 1X TBE by microwaving. Ethidium bromide was then added to a final concentration of 500ng/ml. The agarose was poured in to sealed plastic trays, a comb inserted, and the gel left to set. Samples were run in 1X gel loading buffer. The marker routinely used was PhiX174 DNA/*HaeIII* (Promega). If large nucleic acid molecules (>1kb) or greater resolution was required, NuSieve 3:1 agarose (FMC Bioproducts) was used (prepared as above). After electrophoresis, gels were visualised by UV trans-illumination and a photographic record produced.

7.1.8 Restriction Digests

Digests were performed using Roche or New England Biolabs enzymes and buffers. The restriction enzyme was added to DNA in a 5µl reaction with the appropriate buffer, and incubated at 37°C for 2 hours. The enzymes used for cloning *Dnajc20* in to the pGEX-KG vector for antibody production were *XbaI* at the N-terminal and *HindIII* at the C-terminal. For cloning *Dnajc20* in to the p3xFLAG-CMV™-10 (Sigma) for localisation studies *HindIII* was used to cut the N-terminal, and *EcoRI* was used to cut the C-terminal.

7.1.9 Ligations

Ligations were performed using the pGEX-KG bacterial overexpression vector and the p3xFLAG-CMV™-10 (Sigma) mammalian expression vector. An amplified *Dnajc20* RT-PCR product, using the appropriate primers (Table 7.2), was digested using the appropriate restriction enzymes. The digested product was purified using a PCR purification kit (Qiagen) and was used as an insert. The vectors were digested with the appropriate enzymes and the products separated by electrophoresis on a NuSieve 3:1 gel. The appropriate nucleic acid molecule was purified using a Gel Purification kit (Promega). To maximise efficiency, ligation reactions were set up with the following vector:insert molar ratios; 1:3, 1:1, and 3:1. A vector only control and an insert only control were routinely undertaken. The vector and insert nucleic acid molecules were mixed with DNA ligase and incubated at 4°C for 4 hours. Ligations were confirmed by gel electrophoresis.

7.2 Microbiology

Aseptic technique was used for all procedures involving the growth of bacterial cells. All liquid cultures were grown with vigorous shaking in an orbital shaker (220rpm), and dry cultures were inverted and incubated overnight at 37°C.

7.2.1 Solutions

All reagents were analytical grade and supplied by Sigma, Invitrogen, BDH, Roche, Qiagen and Fisher and stored at 4°C unless otherwise stated. Bacterial growth media stocks were prepared in dated batches and autoclaved by HGU technical service staff.

L-Broth

Tryptone	10.0g
Yeast extract	5.0g
NaCl	10.0g
Glucose	1.0g

dH_2O was added to a final volume of 1 litre.

L-agar

Tryptone	10.0g
Yeast extract	5.0g
NaCl	10.0g
Agar	15.0g

dH_2O was added to a final volume of 1 litre.

Ampicillin, 1000X

100 mg Ampicillin (D-[-]-α-aminobenzylpenicillin sodium salt) was diluted in 2ml ethanol and stored at -20°C. Ampicillin was used at a working concentration 50μg/ml.

7.2.2 Competent cells preparation

Calcium competent BL21 *E. coli* cells were prepared for overexpression of the GST tagged Dnajc20 protein from the pGEX-KG vector. BL21 cells were grown overnight in 100ml of L-broth to saturation. The culture is then centrifuged at 3,200 rpm at 4°C to collect the cells. The cells are re-suspended in 20ml of 0.1M $CaCl_2$ for 15 minutes on ice. The cells are centrifuged at 4°C for 5 minutes at 3,000 rpm. The supernatant is removed, and the cells re-suspended in 2ml of 0.1M $CaCl_2$. The competent cells are kept cold at all times and either used immediately or snap frozen and stored at -70°C.

7.2.3 Transformations

Heat shock transformations were performed by adding 5μl of DNA ligation mix to 95μl of thawed DH5α sub-cloning efficiency *E. coli* (Life Technologies), or calcium competent BL21 cells, in a 1.5ml round bottomed centrifuge tube on ice. The DNA/cell suspension is incubated on ice for 30 minutes, then immersed in a 37°C waterbath for 45 seconds, followed by incubation on ice for 2 minutes. The transformed cells are then suspended in 1ml of L-broth and incubated at 37°C in an orbital shaker for 1 hour.

Using an inoculation loop, the bacteria are spread on to a L-agar plate containing 1X ampicillin (both vectors used are selected by ampicillin resistance). Normally two plates would be used for each transformation, one spread with 50µl and one with 350µl. The L-agar plates are inverted and incubated overnight at 37°C. Control experiments are performed where the original, undigested vector, and untransformed cells are also plated.

7.2.4 Isolation of plasmid DNA

Plasmid DNA was prepared using either the Qiagen miniprep or maxiprep kits. For the extraction of a small amount of plasmid DNA the miniprep kit is used. A single colony is used to inoculate 5ml of L-broth containing 1X ampicillin. The culture is grown at 37°C overnight in an orbital shaker. The following day, plasmid DNA is extracted using the kit, according to the manufacturers instructions. For recovery of large amounts of plasmid DNA a maxiprep is performed. A single colony is used to inoculate 5ml of L-broth and is grown overnight, as before. The following day the culture is stored at 4°C, before being used to inoculate 100ml of L-broth and is grown in an orbital shaker overnight. The following day a maxiprep is performed according to the manufacturers instructions.

7.2.5 Overexpression from the pGEX-KG vector

Transformed BL21 cells are grown overnight in 20ml of L-broth at 37°C in an orbital shaker. The following morning this preculture is added to 400ml of L-broth and grown until the optical density (OD) is between 0.4 and 0.8, approximately 3 hours (1ml of the culture is removed as a –IPTG control sample). 2ml of 20mg/ml IPTG (isopropyl-beta-D-thiogalactopyranoside) is added to the culture to induce expression from the pGEX-KG vector. The IPTG induces expression by inhibiting the *lac* repressor, allowing expression from the *lacZ* promoter. The culture is then grown at 30°C for 4 hours. After these 4 hours 1ml of the sample is removed as a +IPTG control sample. The culture is centrifuged at 4,000rpm for 30 minutes at 4°C and the supernatant removed (the pellet can be stored at -70°C). The pellet is re-suspended in 20ml of GST binding buffer (see 7.3 Protein manipulations). The sample can then be stored, or the protein extracted by sonication.

7.3 Protein manipulation

All reagents and solutions were analytical grade and supplied by Sigma, Invitrogen (Novex), Bio-rad) and Fisher. General lab solutions are prepared by HGU technical service staff, autoclaved, and stored at 4°C unless otherwise stated. All protocols are performed on ice, unless stated otherwise.

7.3.1 Solutions

GST Binding buffer

Tris.HCl pH 7.5 (1M)	5ml
NaCl (5M)	2ml
Glycerol (50%)	20ml
Triton (10%)	10ml
PMSF (100mM)	50µl
Protease cocktail (100X)	10ml

PBS, 10X

NaCl	80g
KCl	2g
Na_2HPO_4	26.8g
KH_2PO_4	2.4g

Dissolve in 800ml dH_2O, adjust the pH to 7.4 with concentrated HCl. Adjust volume to 1 litre and divide in to aliquots for dilution to 1X.

PBT

Add 1ml Tween-20 to 1 litre PBS.

Thrombin protease

Dissolve 500µg in 0.5ml PBS. Aliquot as 50µl working stocks. Snap freeze and store at -70°C.

Modified RIPA buffer, 1X

Tris.HCl pH 7.4	50mM

NP-40	1%
Sodium deoxycholate	0.25%
NaCl	150mM
EDTA	1mM
PMSF	1mM
Protease cocktail	

Separating buffer, 4X

Tris Base	90.75g
SDS (10%)	20ml

Adjust pH to 8.8 with concentrated HCl.

Separating gel, 12%

For 1 minigel

Acrylamide (29:1)	2ml
Separating buffer	1.25ml
dH_2O	1.75ml
APS	50µl
Temed	5µl

Add the APS and Temed immediately before pouring the gel. Pour a layer of dH_2O on top of the gel and leave to set (30 minutes).

Stacking buffer, 4X

Tris Base	12.1g
SDS (8%)	8ml

Adjust pH to 6.8 with concentrated HCl.

Stacking gel

For 1 minigel

Acrylamide (29:1)	0.65ml
Stacking buffer	1.25ml
dH_2O	3.1ml
APS	50µl
Temed	5µl

Add the APS and Temed immediately before pouring the gel. Insert the comb, and leave to set (30 minutes).

SDS Loading buffer, 2X

Tris.HCl (pH 6.8)	1.5ml
SDS (10%)	1.2ml
Glycerol	3ml
β-mercaptoethanol	1.5ml

Adjust volume to 10ml with dH_2O. Make in a fume hood.

Cracking buffer, 1X

NaP (1M)	200µl
SDS (10%)	2ml
Urea	9.6g
Triton (10%)	1ml
Bromophenol blue (1%)	2ml

Adjust volume to 20ml with dH_2O. Immediately prior to use add 50µl β-mercaptoethanol (fume hood).

Running buffer, 5X

Glycine	94g
Tris Base	15.1g
SDS	5g

Dissolve in 800ml dH_2O to dissolve. Adjust volume to 1 litre with dH_2O.

Wet transfer buffer, 1X

Glycine	28.8g
Tris Base	5.8g

Add 1.6 litres dH_2O to dissolve. Add 400ml methanol.

7.3.2 Purification of GST fused Dnajc20

Following overexpression of GST fused Dnajc20 from the pGEX-KG, the protein is extracted by sonication. A 1ml control solubility sample is removed, and centrifuged for 5 minutes at 13,000rpm. The supernatant is mixed with an equal volume of SDS loading buffer and 1ml of cracking buffer is added to the pellet. The samples are separated by SDS-PAGE (7.3.4) along with the –IPTG and +IPTG control samples (7.2.4), and protein visualised by GelCode® Blue Stain Reagent (Pierce) (7.3.7). If the GST fused Dnajc20 protein has been overexpressed and is present in the soluble sample it can be purified. The total sonicated sample is centrifuged for 30 minutes at 4°C. The supernatant is transferred to a fresh 50ml falcon tube. To the sample add 400µl of GS4B sepharose beads (Amersham Biosciences) that are in a 1:1 slurry with PBS. The sample/beads are turned on a wheel for 2 hours at 4°C allowing the GST fused Dnajc20 to bind to the beads. Following binding the supernatant is removed to the top of the beads. The beads are washed 5 times in 4ml of GST binding buffer, with 1minute

centrifugation at 6,500rpm at 4°C between washes. An aliquot of beads can be boiled in SDS loading buffer, and the sample run by SDS-PAGE to confirm purification. The Dnajc20 protein was removed from the beads by thrombin cleavage. Add 14µl of thrombin from working stock to 130µl of PBS. GST binding buffer is removed from the beads and replaced with the diluted thrombin protease. Overnight, the sample is turned on a wheel at room temperature. The Dnajc20 protein is present in the supernatant, and can be confirmed by SDS-PAGE.

7.3.3 Extraction of protein from E10.5 embryos

Immediately following dissection of E10.5 embryos, each embryo is added to a 1.5ml centrifuge tube containing ice-cold modified RIPA buffer. Each sample is kept on ice for 30 minutes. Every 5 to 10 minutes the embryo is pipetted to break up the tissue. After 30 minutes the samples are centrifuged for 15 minutes at 4°C. Following centrifugation the supernatant (containing soluble protein) is transferred to a fresh 1.5ml centrifuge tube (this can be stored at -20°C, but avoid repeated freeze-thaw cycles). To run by either SDS-PAGE or Native-PAGE add an equal volume of the appropriate loading buffer. The pellet can also be kept and Cracking buffer added to investigate the insoluble protein.

7.3.4 SDS-PAGE

The composition of the gel varied, depending upon the protein being detected but routinely a 12% minigel was used to separate proteins according to molecular weight. The gel apparatus was assembled (Bio-rad) and the gel allowed to set at room temperature. The samples were boiled for either 5 minutes in SDS-loading buffer, or 10 minutes in Cracking buffer. Alongside the samples either the Precision Plus Protein™ Kaleidoscope™ protein marker (Bio-rad) or the Prestained protein marker Broad range (P77083, Biolabs) was loaded. Running buffer was added to the gel, and the gel run at 200V for 45-60 minutes. The gel was then either transferred to Hybond-P (Amersham biosciences) for Western blot analysis, or GelCode® Blue Stain Reagent (Pierce) used to detect total protein.

7.3.5 Native-PAGE

Protein from E10.5 embryos was separated by Native-PAGE for investigation of diphthamide on eEF2. Pre-cast 4%-12% Tris-glycine gels (Novex) were used as per the manufacturers instructions. Prior to loading, the protein extract was incubated for 10 minutes at room temperature in 1X native sample buffer (Novex). 10X native-running buffer (Novex) was diluted and added to the gel system (Novex). The samples were run four 4 hours at 125V. The gels were then transferred to Hybond-P for Western blot analysis.

7.3.6 Western blot analysis

A wet blotter was used to transfer proteins onto a Hybond-P filter. The samples were set up in the blotter apparatus following soaking of the pads in transfer buffer, and filling the tank with transfer buffer. From the cathode plate as follows; 1st plastic mesh, 3 wet pads, 1 piece of filter paper cut to the size of the gel and soaked in transfer buffer, the gel washed in transfer buffer (being careful to remove air bubbles), the Hybond-P membrane, soaked in methanol (being careful to remove air bubbles), the second section of filter paper, the remaining wet pads, the 2nd mesh and finally the anode plate and plastic cover. The tank was topped up with transfer buffer and the transfer carried out for 60 minutes.

Once the transfer had taken place, the blot was probed for the protein of interest. Firstly, the hybond-P was blocked in 5% fat free milk in PBT at 4°C overnight (room temperature overnight when probing for the Dnajc20 protein). The primary antibody (α-eEF2, sc-13004 from Santa Cruz, used at a concentration of 1in 500. α-Dnajc20, see 7.3.8, used at a concentration of 1 in 500) was then added to fresh 5% fat free milk in PBT and incubated with the blot for 1 hour at room temperature. Following incubation with the primary antibody, 3 x 15 minute washes were performed with PBT at room temperature. Following completion of the washes the secondary antibody, conjugated with horseradish peroxidase, was then added in fresh 5% fat free milk in PBT (for eEF2, sc-2020 donkey anti-goat, Santa Cruz at 1 in 5000. For Dnajc20, NA934VS anti-rabbit,

Molecular Probes, at 1 in 10,000) for 45 minutes at room temperature. After the incubation, the blot is then washed for 3 x 15 minutes in PBT. Detection of the protein was carried out using ECL-plus (Amersham Biosciences) as described by the manufacturer. Bands were visualised by incubating for 1 minute with Kodak BioMax XAR films in and developed using an X-ray processor.

7.3.7 Visualisation of total protein

To visualise total protein GelCode® Blue Stain Reagent (Pierce) was used according to the manufacturers instructions. Briefly, following SDS-PAGE the gel is placed in a microwave proof tray with 100ml of ultrapure water and heated for 90 seconds. The water is replaced with 100ml of fresh ultrapure water and microwaved again for 90 seconds. The water is replaced with fresh ultrapure water and washed for 5 minutes. The water is then replaced with 50ml of GelCode® Blue Stain Reagent (Pierce) to completely cover the gel, and microwaved for 1 minute. After the microwave step, incubate for 5 minutes at room temperature. Following the staining the gel is washed for at least 10 minutes in 200ml of ultrapure water. Longer washes, regularly changing the water, can reduce background. The gel then can be dried using a vacuum dryer, and stored indefinitely.

7.3.8 Production of an antibody against Dnajc20

The full-length *Dnajc20* gene was amplified by RT-PCR and cloned into the pGEX-KG vector (7.1.6 to 7.1.9). A GST fused Dnajc20 construct was then produced by overexpression in BL21 *E. coli* (7.2.4) and purified by GST pull-down and thrombin cleavage (7.3.2). The full-length Dnajc20 protein was the sent on ice to Diagnostics Scotland for contract immunisation. The anti-Dnajc20 antibody was raised in a single rabbit. A pre-immunisation bleed was taken, followed by bleeds at month 1 and month 2 following immunisation. At month 3 the rabbit was exsanguinated. This final bleed was used directly in Western analysis against the purified Dnajc20 protein and found to detect a single band corresponding to the expected size. This sera detects multiple bands by Western analysis against total protein from E10.5 embryos, including a band of around 17kDa, which corresponds to the expected size of the Dnajc20 protein. This band is absent in Western analysis performed protein from E10.5 *Dnajc20* homozygous mutant embryos.

7.4 Mammalian cell culture and immunofluorescence

All reagents were produced by HGU technical service staff and stored at 4°C unless otherwise stated. All tissue culture was performed in tissue culture hoods.

7.4.1 Reagents and solutions

Tissue culture media

Dulbecco's media (DMEM)	500ml
Fetal Calf Serum (FCS)	50ml
Penicillin/streptomycin	5ml
Glutamine	5ml

Heat in a waterbath at 37°C for 30 minutes prior to use.

Trypsin

Versine	10ml
Trypsin	10ml

Mix together and store at -20°C until needed. Heat in a waterbath at 37°C for 30 minutes prior to use.

4% PFA

Paraformaldehyde	2g
PBS	50ml

Heat PBS for 1 minute in a microwave. Add to PFA in a fume hood and allow to dissolve. Prior to use the 4% PFA should be ice cold.

7.4.2 Cell culture and splitting

NIH-3T3 cells were routinely used. Cell lines were maintained at 37°C with 5% CO_2 in tissue culture media. Cells were grown in 10cm petri dishes in 10ml tissue culture media to near confluence. To split the cells, the media is poured off, followed by two washes in PBS (37°C). 1ml of trypsin is then added, and incubated for 5 minutes at 37°C. Gentle agitation dislodges the cells, and 9ml fresh tissue culture media added. Repeated pipetting separates toe cells. 1ml of this is added to a new 10cm petri dish containing 9ml fresh tissue culture media.

7.4.3 Transfection of NIH-3T3

NIH-3T3 cells are seeded onto coverslips in 6-well plates, and grown to >90% confluence. For each well, 5µl Lipofectamine 2000 (Invitrogen) is added to 200µl of Opti-MEM (Invitrogen), gently mixed, and incubated for 5 minutes at room temperature. 2µg of the DNA for transfection is incubated with 200µl of Opti-MEM and incubated for 5 minutes. Following incubation, the two mixtures are combined, mixed gently, and incubated at room temperature for a further 20 minutes. The mixture is then added to the well by drop-wise addition, while the 6-well plate is rocked gently from side-to-side. The cells are then incubated for 44 to 48 hours. The tissue culture media is not replaced.

7.4.4 Immunofluorescence on fixed cells

Following transfection of cells and growth on coverslips, 2 washes in ice-cold PBS were performed. After this ice-cold 4% PFA was added, and the cells fixed on ice for 10 minutes followed by 2 x 5 minute washes in PBT washes in P. Cells were then blocked for 1 hour at room temperature in PBS containing azide and BSA (bovine serum albumin). Following blocking, the primary antibody was added to fresh PBS-azide-BSA to a total volume of 400µl (either anti-FLAG M2 monoclonal F-3165, Sigma, at a concentration of 1 in 200, or C-Myc (A-14) sc-789, Santa cruz, at a concentration of 1 in 100) and incubated for 1 hour at room temperature on a rack. After incubation, 3 x 5 minute washes in PBT were carried out. The secondary antibodies or stains were then added to fresh PBS-azide-BSA to a total volume of 400µl (anti-Mouse Texas Red, 715-075-150, Stratech, at a concentration of 1 in 400. anti-mouse Alexa 488, 715-095-150, Stratatech. anti-Rabbit Alexa 488, A11034, Molecular probes, at a concentration of 1 in 400. Phalloidin R415, Molecular probes, at a concentration of 1 in 200). Incubation was performed on racks at room temperature in the dark. Following incubation, 3 x 5 minute washes in PBT were performed. Coverslips were mounted on to slides in Vectashield (Vector). Coverslips were sealed using vulcanising solution (REMA Tip Top), and slides stored at 4°C in the dark.

7.5 Computational methods

7.5.1 Programs and databases

BLAST (version 2.1) Altschul *et al* 1997.
BLAST was performed at the following databases
http://www.hgmp.mrc.ac.uk
http://www.ncbi.nlm.nih.gov/BLAST/
http://www.ensembl.org/Multi/blastview

ClustalW (version 1.74) Thompson *et al* 1994[168]
Sequence alignment program.

Ensembl (Multiple builds) Clamp *et al* 2003[169]
INTERPRO Apweiler *et al* 2001
http://www.ensembl.org/Mus_musculus/protview?db=core;peptide=ENSMUSP00000030118

GeneDoc Nicholas *et al* (unpublished)[170]

Mouse Genome Informatics (MGI)
http://www.informatics.jax.org/

Pfam Bateman A *et al* (2000)[171]
www.sanger.ac.uk/Software/Pfam/

PSORT II Nakai K and Horton P (1999)[52]
http://psort.nibb.ac.jp/form2.html

SMART Schultz *et al* 1998[172]. Letunic *et al* 2004[173]
http://smart.embl-heidelberg.de/

7.5.2 Identification of mouse J-proteins

Protein sequences, annotated as J-proteins, were retrieved from the MGI database (according to name) and the Ensembl database (according to the presence of the INTERPRO IPR001623 entry). The proteins were confirmed in the Ensembl database, and the J-protein omitted if it did not have a proper gene structure. BLAST searches were performed with each of the J-proteins to identify any proteins represented more than once in the dataset. This J-protein dataset was then entered in to the SMART database to confirm the presence of a defined J-domain based upon the Pfam annotation. This resulted in the final dataset of mouse J-proteins. Initial characterisation of domain structure was performed on the entire dataset through SMART analysis. Novel J-proteins were subjected to PSORT II analysis for characterisation.

7.5.3 Identification of sequence orthologues of Dnajc20

The protein sequence of the mouse Dnajc20 protein was used in 3 iteration PSI- and PHI-BLAST searches to identify protein orthologues. The J-domain containing sequence was masked to reduce results that identify other J-proteins. If a sequence showed homology to the C-terminal region of Dnajc20 it was investigated for the

presence of an N-terminal J-domain. This analysis also identified the presence of the small zf-CSL proteins. Many of the proteins have subsequently been annotated as zf-CSL proteins by Pfam.

7.5.4 Alignment of protein sequence

Protein sequence alignments were created using ClustalW, with an MSF output. The alignments were visualised using GeneDoc.

7.6 Mouse Husbandry

All animals were maintained in the BRF facility by University of Edinburgh staff and Lisa Mckie, a member of our group. Wild-type mice were bred in-house. Embryos from all experiments were generated from timed matings, with the morning of vaginal plug detection being counted as embryonic day 0.5. Tissue for genotyping was obtained by ear-clipping following weaning.

7.7 Microscopy

7.7.1 Wholemount microscopy

Dissected embryos were imaged with a Leica Stereo MZFLIII stereofluorescence microscope (Leica Microsystems Milton Keynes, UK). Embryos were illuminated from above by a separate light source. Images of whole embryos under PBS were captured from above with a Photometrics CoolSnap colour CCD camera (Roper Scientific, Arizona) controlled scripts written for IPLab Spectrum (Scanalytics, VA).

7.7.2 Fluorescence imaging of cells

Slides for immunofluorescence experiments were examined using a Zeiss Axioplan II fluorescence microscope with 100 watt mercury bulbs and equipped with a triple band-pass filter (Chroma # 83000). Images were collected using IPLAB v3.6 (Scanalystics, VA).

Chapter 8

References

Reference List

1. Lander,E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

2. Venter,J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304-+ (2001).

3. Goffeau,A. *et al.* Life with 6000 genes. *Science* **274**, 546-& (1996).

4. [Anon]. The C-elegans sequencing consortium (vol 282, pg 2012, 1998). *Science* **283**, 2103 (1999).

5. Adams,M.D. *et al.* The genome sequence of Drosophila melanogaster. *Science* **287**, 2185-2195 (2000).

6. Waterston,R.H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562 (2002).

7. Winzeler,E.A. *et al.* Functional characterization of the S-cerevisiae genome by gene deletion and parallel analysis. *Science* **285**, 901-906 (1999).

8. Hartwell,L.H., Culotti,J. & Reid,B. Genetic Control of Cell-Division Cycle in Yeast .1. Detection of Mutants. *Proceedings of the National Academy of Sciences of the United States of America* **66**, 352-& (1970).

9. Hartwell,L.H. & Weinert,T.A. Checkpoints - Controls That Ensure the Order of Cell-Cycle Events. *Science* **246**, 629-634 (1989).

10. Nurse,P., Thuriaux,P. & Nasmyth,K. Genetic-Control of Cell-Division Cycle in Fission Yeast Schizosaccharomyces-Pombe. *Molecular & General Genetics* **146**, 167-178 (1976).

11. Daum,G. *et al.* Systematic analysis of yeast strains with possible defects in lipid metabolism. *Yeast* **15**, 601-614 (1999).

12. Steinmetz,L.M. *et al.* Systematic screen for human disease genes in yeast. *Nature Genetics* **31**, 400-404 (2002).

13. Chang,M., Bellaoui,M., Boone,C. & Brown,G.W. A genome-wide screen for methyl methanesulfonate-sensitive mutants reveals genes required for S phase progression in the presence of DNA damage. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16934-16939 (2002).

14. Nussleinvolhard,C. & Wieschaus,E. Mutations Affecting Segment Number and Polarity in Drosophila. *Nature* **287**, 795-801 (1980).

15. Siegfried,E., Wilder,E.L. & Perrimon,N. Components of Wingless Signaling in Drosophila. *Nature* **367**, 76-80 (1994).

16. Muqit,M.M.K. & Feany,M.B. Modelling neurodegenerative diseases in Drosophila: a fruitful approach? *Nature Reviews Neuroscience* **3**, 237-243 (2002).

17. Maeda,I., Kohara,Y., Yamamoto,M. & Sugimoto,A. Large-scale analysis of gene function in Caenorhabditis elegans by high-throughput RNAi. *Current Biology* **11**, 171-176 (2001).

18. Hariharan,I.K. & Haber,D.A. Yeast, flies, worms, and fish in the study of human disease. *New England Journal of Medicine* **348**, 2457-2463 (2003).

19. Ashburner,M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25-29 (2000).

20. Manoukian S,C. *et al.* Bilateral preaxial polydactyly in a WAGR syndrome patient. Am J Med Genet A. 134(4), 426-429. 2005.

21. Bremond-Gignac D,G.*et al.* Three patients with hallucal polydactyly and WAGR syndrome, including discordant expression of Wilms tumor in MZ twins. Am J Med Genet A 134(4), 422-425. 2005.

22. Kroczynska,B. & Blond,S.Y. Cloning and characterization of a new soluble murine J-domain protein that stimulates BiP, Hsc70 and DnaK ATPase activity with different efficiencies. *Gene* **273**, 267-274 (2001).

23. Liu,S., Milne,G.T., Kuremsky,J.G., Fink,G.R. & Leppla,S.H. Identification of the proteins required for biosynthesis of diphthamide, the target of bacterial ADP-ribosylating toxins on translation elongation factor 2. *Mol. Cell Biol.* **24**, 9487-9497 (2004).

24. Dunlop,P.C. & Bodley,J.W. Biosynthetic labeling of diphthamide in Saccharomyces cerevisiae. *J. Biol. Chem.* **258**, 4754-4758 (1983).

25. Van Ness,B.G., Howard,J.B. & Bodley,J.W. ADP-ribosylation of elongation factor 2 by diphtheria toxin. NMR spectra and proposed structures of ribosyl-diphthamide and its hydrolysis products. *J. Biol. Chem.* **255**, 10710-10716 (1980).

26. Chen,C.M. & Behringer,R.R. Ovca1 regulates cell proliferation, embryonic development, and tumorigenesis. *Genes & Development* **18**, 320-332 (2004).

27. Shibuya,T. & Morimoto,K. A Review of the Genotoxicity of 1-Ethyl-1-Nitrosourea. *Mutation Research* **297**, 3-38 (1993).

28. Justice,M.J., Noveroske,J.K., Weber,J.S., Zheng,B.H. & Bradley,A. Mouse ENU mutagenesis. *Human Molecular Genetics* **8**, 1955-1963 (1999).

29. Noveroske,J.K., Weber,J.S. & Justice,M.J. The mutagenic action of N-ethyl-N-nitrosourea in the mouse. *Mammalian Genome* **11**, 478-483 (2000).

30. Rinchik,E.M., Carpenter,D.A. & Selby,P.B. A Strategy for Fine-Structure Functional-Analysis of A 6-Centimorgan to 11-Centimorgan Region of Mouse Chromosome-7 by High-Efficiency Mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 896-900 (1990).

31. Quwailid,M.M. *et al.* A gene-driven ENU-based approach to generating an allelic series in any gene. *Mammalian Genome* **15**, 585-591 (2004).

32. de Angelis,M.H. *et al.* Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nature Genetics* **25**, 444-447 (2000).

33. Nolan,P.M. *et al.* A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat. Genet.* **25**, 440-443 (2000).

34. Herron,B.J. *et al.* Efficient generation and mapping of recessive developmental mutations using ENU mutagenesis. *Nature Genetics* **30**, 185-189 (2002).

35. Hogan,B.L.M., Horsborough,G. & Cohen,J. Small Eye (Sey) Is A Homozygous Lethal Mutation Affecting the Development of Both Lens and Nasal Placodes. *Genetical Research* **47**, 220 (1986).

36. Glaser,T. & Housman,D.E. The Small Eye Mouse (Sey), An Animal-Model Aniridia (An2). *Cytogenetics and Cell Genetics* **51**, 1005 (1989).

37. Glaser,T., Lane,J. & Housman,D. A Mouse Model of the Aniridia-Wilms Tumor Deletion Syndrome. *Science* **250**, 823-827 (1990).

38. Hill,R.E. *et al.* Mouse Small Eye Results from Mutations in A Paired-Like Homeobox-Containing Gene. *Nature* **354**, 522-525 (1991).

39. Buckler,A.J., Pelletier,J., Haber,D.A., Glaser,T. & Housman,D.E. Isolation, Characterization, and Expression of the Murine Wilms-Tumor Gene (Wt1)

During Kidney Development. *Molecular and Cellular Biology* **11**, 1707-1712 (1991).

40. Roberts CR. **Small-eyes**, a new dominant mutant in the mouse. Genet.Res. 9, 121-122. 1967.

41. Favor,J. & Neuhauser-Klaus,A. Saturation mutagenesis for dominant eye morphological defects in the mouse Mus musculus. *Mammalian Genome* **11**, 520-525 (2000).

42. Favor,J. *et al.* Molecular characterization of Pax6(2Neu) through Pax6(10Neu): An extension of the Pax6 allelic series and the identification of two possible hypomorph alleles in the mouse Mus musculus. *Genetics* **159**, 1689-1700 (2001).

43. Theiler,K., Varnum,D.S. & Stevens,L.C. Development of Dickies Small Eye, A Mutation in House Mouse. *Anatomy and Embryology* **155**, 81-86 (1978).

44. Lyon,M.F., Phillips,R.J.S. & Fisher,G. Dose-Response Curves for Radiation-Induced Gene-Mutations in Mouse Oocytes and Their Interpretation. *Mutation Research* **63**, 161-173 (1979).

45. Hogan,B.L.M. *et al.* Small Eyes (Sey) - A Homozygous Lethal Mutation on Chromosome 2 Which Affects the Differentiation of Both Lens and Nasal Placodes in the Mouse. *Journal of Embryology and Experimental Morphology* **97**, 95-110 (1986).

46. Kreidberg,J.A. *et al.* Wt-1 Is Required for Early Kidney Development. *Cell* **74**, 679-691 (1993).

47. Kent,J. *et al.* The reticulocalbin gene maps to the WAGR region in human and to the small eye Harwell deletion in mouse. *Genomics* **42**, 260-267 (1997).

48. Kleinjan,D.A., Seawright,A., Elgar,G. & van Heyningen,V. Characterization of a novel gene adjacent to PAX6, revealing synteny conservation with functional significance. *Mammalian Genome* **13**, 102-107 (2002).

49. Webb,T.R. Mouse mutagenesis screens: the identification of candidate genes using computational analysis and the characterisation of a mutation in the *Mitf* gene. Unpublished . 2002. (Masters Thesis)

50. Zeng,L. *et al.* Identification of a novel human doublecortin-domain-containing gene (DCDC1) expressed mainly in testis. *Journal of Human Genetics* **48**, 393-396 (2003).

51. Moore,A.W. *et al.* YAC transgenic analysis reveals Wilms' Tumour 1 gene activity in the proliferating coelomic epithelium, developing diaphragm and limb. *Mechanisms of Development* **79**, 169-184 (1998).

52. Nakai,K. & Sakamoto,H. Construction of A Novel Database Containing Aberrant Splicing Mutations of Mammalian Genes. *Gene* **141**, 171-177 (1994).

53. Walsh,P., Bursac,D., Law,Y.C., Cyr,D. & Lithgow,T. The J-protein family: modulating protein assembly, disassembly and translocation. *EMBO Rep.* **5**, 567-571 (2004).

54. Kelley,W.L. The J-domain family and the recruitment of chaperone power. *Trends in Biochemical Sciences* **23**, 222-227 (1998).

55. Tavaria,M., Gabriele,T., Kola,I. & Anderson,R.L. A hitchhiker's guide to the human Hsp70 family. *Cell Stress & Chaperones* **1**, 23-28 (1996).

56. Cheetham,M.E. & Caplan,A.J. Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. *Cell Stress. Chaperones.* **3**, 28-36 (1998).

57. Mayer,M.P. *et al.* Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nature Structural Biology* **7**, 586-593 (2000).

58. Bukau,B. & Horwich,A.L. The Hsp70 and Hsp60 chaperone machines. *Cell* **92**, 351-366 (1998).

59. Flynn,G.C., Pohl,J., Flocco,M.T. & Rothman,J.E. Peptide-Binding Specificity of the Molecular Chaperone Bip. *Nature* **353**, 726-730 (1991).

60. Schmid,D., Baici,A., Gehring,H. & Christen,P. Kinetics of Molecular Chaperone Action. *Science* **263**, 971-973 (1994).

61. McCarty,J.S., Buchberger,A., Reinstein,J. & Bukau,B. The Role of Aip in the Functional Cycle of the Dnak Chaperone System. *Journal of Molecular Biology* **249**, 126-137 (1995).

62. Theyssen,H., Schuster,H.P., Packschies,L., Bukau,B. & Reinstein,J. The second step of ATP binding to DnaK induces peptide release. *Journal of Molecular Biology* **263**, 657-670 (1996).

63. Mayer,M.P. & Bukau,B. Hsp70 chaperone systems: Diversity of cellular functions and mechanism of action. *Biological Chemistry* **379**, 261-268 (1998).

64. Sunshine,M., Feiss,M., Stuart,J. & Yochem,J. New Host Gene (Gropc) Necessary for Lambda-Dna Replication. *Molecular & General Genetics* **151**, 27-34 (1977).

65. Yochem,J. *et al.* Genetic-Analysis of 2 Genes, Dnaj and Dnak, Necessary for Escherichia-Coli and Bacteriophage-Lambda Dna-Replication. *Molecular & General Genetics* **164**, 9-14 (1978).

66. Georgopoulos,C.P., Lundquist-Heil,A., Yochem,J. & Feiss,M. Identification of the E. coli dnaJ gene product. *Mol. Gen. Genet.* **178**, 583-588 (1980).

67. Itikawa,H. & Ryu,J.I. Isolation and Characterization of A Temperature-Sensitive Dnak Mutant of Escherichia-Coli-B. *Journal of Bacteriology* **138**, 339-344 (1979).

68. Karzai,A.W. & McMacken,R. A bipartite signaling mechanism involved in DnaJ-mediated activation of the Escherichia coli DnaK protein. *Journal of Biological Chemistry* **271**, 11236-11246 (1996).

69. Linke,K., Wolfram,T., Bussemer,J. & Jakob,U. The roles of the two zinc binding sites in DnaJ. *J. Biol. Chem.* **278**, 44457-44466 (2003).

70. Frazier,A.E. *et al.* Pam16 has an essential role in the mitochondrial protein import motor. *Nature Structural & Molecular Biology* **11**, 226-233 (2004).

71. Ohtsuka,K. & Hata,M. Mammalian HSP40/DNAJ homologs: cloning of novel cDNAs and a proposal for their classification and nomenclature. *Cell Stress & Chaperones* **5**, 98-112 (2000).

72. Kanazawa,M., Terada,K., Kato,S. & Mori,M. HSDJ, a human homolog of DnaJ, is farnesylated and is involved in protein import into mitochondria. *J. Biochem. (Tokyo)* **121**, 890-895 (1997).

73. Terada,K. & Mori,M. Human DnaJ homologs dj2 and dj3, and bag-1 are positive cochaperones of hsc70. *J. Biol. Chem.* **275**, 24728-24734 (2000).

74. Cheetham,M.E., Jackson,A.P. & Anderton,B.H. Regulation of 70-Kda Heat-Shock-Protein Atpase Activity and Substrate-Binding by Human Dnaj-Like Proteins, Hsj1A and Hsj1B. *European Journal of Biochemistry* **226**, 99-107 (1994).

75. Chapple,J.P. & Cheetham,M.E. The chaperone environment at the cytoplasmic face of the endoplasmic reticulum can modulate rhodopsin processing and inclusion formation. *Journal of Biological Chemistry* **278**, 19087-19094 (2003).

76. Chapple,J.P., van der,S.J., Poopalasundaram,S. & Cheetham,M.E. Neuronal DnaJ proteins HSJ1a and HSJ1b: a role in linking the Hsp70 chaperone machine to the ubiquitin-proteasome system? *Biochem. Soc. Trans.* **32**, 640-642 (2004).

77. Ungewickell,E. *et al.* Role of auxilin in uncoating clathrin-coated vesicles. *Nature* **378**, 632-635 (1995).

78. Holstein,S.E., Ungewickell,H. & Ungewickell,E. Mechanism of clathrin basket dissociation: separate functions of protein domains of the DnaJ homologue auxilin. *J. Cell Biol.* **135**, 925-937 (1996).

79. Jiang,R.F., Greener,T., Barouch,W., Greene,L. & Eisenberg,E. Interaction of auxilin with the molecular chaperone, Hsc70. *J. Biol. Chem.* **272**, 6141-6145 (1997).

80. Gruschus,J.M. *et al.* Structure of the functional fragment of auxilin required for catalytic uncoating of clathrin-coated vesicles. *Biochemistry* **43**, 3111-3119 (2004).

81. Greener,T., Nojima,H., Eisenberg,E. & Greene,L.E. Cyclin G associated kinase (GAK) has kinase and auxilin domains that are active in vitro. *Faseb Journal* **12**, A1419 (1998).

82. Greener,T., Zhao,X.H., Nojima,H., Eisenberg,E. & Greene,L.E. Role of cyclin G-associated kinase in uncoating clathrin-coated vesicles from non-neuronal cells. *Journal of Biological Chemistry* **275**, 1365-1370 (2000).

83. Greener,T., Eisenberg,E. & Greene,L. Auxilin and GAK have a unique clathrin binding domain that enables C-terminal 20 kDa recombinant fragments of these proteins to support uncoating of clathrin baskets by Hsc70 but not to induce clathrin polymerization. *Faseb Journal* **14**, A1493 (2000).

84. Chang,H.C., Hull,M. & Mellman,I. The J-domain protein Rme-8 interacts with Hsc70 to control clathrin-dependent endocytosis in Drosophila. *Journal of Cell Biology* **164**, 1055-1064 (2004).

85. Braun,J.E.A., Wilbanks,S.M. & Scheller,R.H. The cysteine string secretory vesicle protein activates hsc70 ATPase. *Journal of Biological Chemistry* **271**, 25989-25993 (1996).

86. Gautschi,M. *et al.* RAC, a stable ribosome-associated complex in yeast formed by the DnaK-DnaJ homologs Ssz1p and zuotin. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 3762-3767 (2001).

87. Minami,Y., Hohfeld,J., Ohtsuka,K. & Hartl,F.U. Regulation of the heat-shock protein 70 reaction cycle by the mammalian DnaJ homolog, Hsp40. *J. Biol. Chem.* **271**, 19617-19624 (1996).

88. Luke,M.M., Sutton,A. & Arndt,K.T. Characterization of SIS1, a Saccharomyces cerevisiae homologue of bacterial dnaJ proteins. *J. Cell Biol.* **114**, 623-638 (1991).

89. Baker,R.K. *et al.* In vitro preselection of gene-trapped embryonic stem cell clones for characterizing novel developmentally regulated genes in the mouse. *Developmental Biology* **185**, 201-214 (1997).

90. Hunter,P.J., Swanson,B.J., Haendel,M.A., Lyons,G.E. & Cross,J.C. Mrj encodes a DnaJ-related co-chaperone that is essential for murine placental development. *Development* **126**, 1247-1258 (1999).

91. Terada,K. *et al.* A type I DnaJ homolog, DjA1, regulates androgen receptor signaling and spermatogenesis. *Embo Journal* **24**, 611-622 (2005).

92. Kurzikdumke,U., Phannavong,B., Gundacker,D. & Gateff,E. Genetic, Cytogenetic and Developmental Analysis of the Drosophila-Melanogaster Tumor Suppressor Gene Lethal(2)Tumorous Imaginal Disks (1(2)Tid). *Differentiation* **51**, 91-104 (1992).

93. Lo,J.F. *et al.* Tid1, a cochaperone of the heat shock 70 protein and the mammalian counterpart of the Drosophila tumor suppressor l(2)tid, is critical for early embryonic development and cell survival. *Mol. Cell Biol.* **24**, 2226-2236 (2004).

94. Fernandez-Chacon,R. *et al.* The synaptic vesicle protein CSP alpha prevents presynaptic degeneration. *Neuron* **42**, 237-251 (2004).

95. Ladiges,W.C. *et al.* Pancreatic beta-cell failure and diabetes in mice with a deletion mutation of the endoplasmic reticulum molecular chaperone gene P58(IPK). *Diabetes* **54**, 1074-1081 (2005).

96. Cheetham,M.E., Jackson,A.P. & Anderton,B.H. Regulation of 70-kDa heat-shock-protein ATPase activity and substrate binding by human DnaJ-like proteins, HSJ1a and HSJ1b. *Eur. J. Biochem.* **226**, 99-107 (1994).

97. Evans,G.J., Morgan,A. & Burgoyne,R.D. Tying everything together: the multiple roles of cysteine string protein (CSP) in regulated exocytosis. *Traffic.* **4**, 653-659 (2003).

98. Yang W,H.J. Genomic analysis of G protein gamma subunits in human and mouse - The relationship between conserved gene structure and G protein betagamma dimer formation. Cell signal . 2005.

99. Chen,J. *et al.* Cloning and characterization of a novel human cDNA encoding a J-domain protein (DNAJA5) from the fetal brain. *Int. J. Mol. Med.* **13**, 735-740 (2004).

100. Merla,G., Ucla,C., Guipponi,M. & Reymond,A. Identification of additional transcripts in the Williams-Beuren syndrome critical region. *Hum. Genet.* **110**, 429-438 (2002).

101. Gitton,Y. *et al.* A gene expression map of human chromosome 21 orthologues in the mouse. *Nature* **420**, 586-590 (2002).

102. Behr,M., Riedel,D. & Schuh,R. The claudin-like Megatrachea is essential in septate junctions for the epithelial barrier function in Drosophila. *Developmental Cell* **5**, 611-620 (2003).

103. Hennessy,F., Cheetham,M.E., Dirr,H.W. & Blatch,G.L. Analysis of the levels of conservation of the J domain among the various types of DnaJ-like proteins. *Cell Stress. Chaperones.* **5**, 347-358 (2000).

104. Yan,W., Gale,M.J., Jr., Tan,S.L. & Katze,M.G. Inactivation of the PKR protein kinase and stimulation of mRNA translation by the cellular co-chaperone P58(IPK) does not require J domain function. *Biochemistry* **41**, 4938-4945 (2002).

105. Hennessy,F., Nicoll,W.S., Zimmermann,R., Cheetham,M.E. & Blatch,G.L. Not all J domains are created equal: Implications for the specificity of Hsp40-Hsp70 interactions. *Protein Science* **14**, 1697-1709 (2005).

106. Fichtner,L. & Schaffrath,R. KTI11 and KTI13, Saccharomyces cerevisiae genes controlling sensitivity to G1 arrest induced by Kluyveromyces lactis zymocin. *Molecular Microbiology* **44**, 865-875 (2002).

107. Frohloff,F., Fichtner,L., Jablonowski,D., Breunig,K.D. & Schaffrath,R. Saccharomyces cerevisiae Elongator mutations confer resistance to the Kluyveromyces lactis zymocin. *Embo Journal* **20**, 1993-2003 (2001).

108. Jablonowski,D., Frohloff,F., Fichtner,L., Stark,M.J.R. & Schaffrath,R. Kluyveromyces lactis zymocin mode of action is linked to RNA polymerase II function via Elongator. *Molecular Microbiology* **42**, 1095-1105 (2001).

109. Kleinjan,D.A., Seawright,A., Elgar,G. & van Heyningen,V. Characterization of a novel gene adjacent to PAX6, revealing synteny conservation with functional significance. *Mammalian Genome* **13**, 102-107 (2002).

110. Otero,G. *et al.* Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Molecular Cell* **3**, 109-118 (1999).

111. Wittschieben,B.O. *et al.* A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Molecular Cell* **4**, 123-128 (1999).

112. Kim,J.H., Lane,W.S. & Reinberg,D. Human Elongator facilitates RNA polymerase II transcription through chromatin. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 1241-1246 (2002).

113. Hawkes,N.A. *et al.* Purification and characterization of the human Elongator complex. *Journal of Biological Chemistry* **277**, 3047-3052 (2002).

114. Slaugenhaupt,S.A. & Gusella,J.F. Familial dysautonomia. *Current Opinion in Genetics & Development* **12**, 307-311 (2002).

115. Mezey,E. *et al.* Of splice and men: what does the distribution of IKAP mRNA in the rat tell us about the pathogenesis of familial dysautonomia? *Brain Research* **983**, 209-214 (2003).

116. Fichtner,L. *et al.* Elongator's toxin-target (TOT) function is nuclear localization sequence dependent and suppressed by post-translational modification. *Molecular Microbiology* **49**, 1297-1307 (2003).

117. Pokholok,D.K., Hannett,N.M. & Young,R.A. Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Molecular Cell* **9**, 799-809 (2002).

118. Krogan,N.J. *et al.* RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomnics approach. *Molecular and Cellular Biology* **22**, 6979-6992 (2002).

119. Rahl,P.B., Chen,C.Z. & Collins,R.N. Elp1p, the yeast homolog of the FD disease syndrome protein, negatively regulates exocytosis independently of transcriptional elongation. *Molecular Cell* **17**, 841-853 (2005).

120. Sjolinder,M., Uhlmann,J. & Ponstingl,H. Characterisation of an evolutionary conserved protein interacting with the putative guanine nucleotide exchange factor DelGEF and modulating secretion. *Experimental Cell Research* **294**, 68-76 (2004).

121. Sjolinder,M., Uhlmann,J. & Ponstingl,H. DelGEF, a homologue of the Ran guanine nucleotide exchange factor RanGEF, binds to the exocyst component Sec5 and modulates secretion. *Febs Letters* **532**, 211-215 (2002).

122. Liu,S.H. & Leppla,S.H. Retroviral insertional mutagenesis identifies a small protein required for synthesis of diphthamide, the target of bacterial ADP-ribosylating toxins. *Molecular Cell* **12**, 603-613 (2003).

123. Liu,S.H., Milne,G.T., Kuremsky,J.G., Fink,G.R. & Leppla,S.H. Identification of the proteins required for biosynthesis of diphthamide, the target of bacterial ADP-ribosylating toxins on translation elongation factor 2. *Molecular and Cellular Biology* **24**, 9487-9497 (2004).

124. Van Ness,B.G., Howard,J.B. & Bodley,J.W. ADP-ribosylation of elongation factor 2 by diphtheria toxin. Isolation and properties of the novel ribosyl-amino acid and its hydrolysis products. *J. Biol. Chem.* **255**, 10717-10720 (1980).

125. Moehring,J.M., Moehring,T.J. & Danley,D.E. Posttranslational Modification of Elongation Factor-Ii in Diphtheria-Toxin-Resistant Mutants of Cho-K1 Cells. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **77**, 1010-1014 (1980).

126. Moehring,T.J., Danley,D.E. & Moehring,J.M. In vitro biosynthesis of diphthamide, studied with mutant Chinese hamster ovary cells resistant to diphtheria toxin. *Mol. Cell Biol.* **4**, 642-650 (1984).

127. Chen,J.Y.C., Bodley,J.W. & Livingston,D.M. Diphtheria Toxin-Resistant Mutants of Saccharomyces-Cerevisiae. *Molecular and Cellular Biology* **5**, 3357-3360 (1985).

128. Collier,R.J. Understanding the mode of action of diphtheria toxin: a perspective on progress during the 20th century. *Toxicon* **39**, 1793-1803 (2001).

129. Oppenheimer,N.J. & Bodley,J.W. Diphtheria toxin. Site and configuration of ADP-ribosylation of diphthamide in elongation factor 2. *J. Biol. Chem.* **256**, 8579-8581 (1981).

130. Pappenheimer,A.M., Jr., Dunlop,P.C., Adolph,K.W. & Bodley,J.W. Occurrence of diphthamide in archaebacteria. *J. Bacteriol.* **153**, 1342-1347 (1983).

131. Chen,J.Y. & Bodley,J.W. Biosynthesis of diphthamide in Saccharomyces cerevisiae. Partial purification and characterization of a specific S-adenosylmethionine:elongation factor 2 methyltransferase. *J. Biol. Chem.* **263**, 11692-11696 (1988).

132. Mattheakis,L.C., Sor,F. & Collier,R.J. Diphthamide synthesis in Saccharomyces cerevisiae: structure of the DPH2 gene. *Gene* **132**, 149-154 (1993).

133. Liu,S., Milne,G.T., Kuremsky,J.G., Fink,G.R. & Leppla,S.H. Identification of the proteins required for biosynthesis of diphthamide, the target of bacterial ADP-ribosylating toxins on translation elongation factor 2. *Mol. Cell Biol.* **24**, 9487-9497 (2004).

134. Mattheakis,L.C., Shen,W.H. & Collier,R.J. DPH5, a methyltransferase gene required for diphthamide biosynthesis in Saccharomyces cerevisiae. *Mol. Cell Biol.* **12**, 4026-4037 (1992).

135. Liu,S. & Leppla,S.H. Retroviral insertional mutagenesis identifies a small protein required for synthesis of diphthamide, the target of bacterial ADP-ribosylating toxins. *Mol. Cell* **12**, 603-613 (2003).

136. Fichtner,L. *et al.* Elongator's toxin-target (TOT) function is nuclear localization sequence dependent and suppressed by post-translational modification. *Molecular Microbiology* **49**, 1297-1307 (2003).

137. Chen,C.M. & Behringer,R.R. Cloning, structure, and expression of the mouse Ovca1 gene. *Biochemical and Biophysical Research Communications* **286**, 1019-1026 (2001).

138. Gavrilova,L.P. *et al.* Immunofluorescent Localization of Protein-Synthesis Components in Mouse Embryo Fibroblasts. *Cell Biology International Reports* **11**, 745-753 (1987).

139. Shestakova,E.A., Motuz,L.P., Minin,A.A., Gelfand,V.I. & Gavrilova,L.P. Some of Eukaryotic Elongation Factor-Ii Is Colocalized with Actin Microfilament Bundles in Mouse Embryo Fibroblasts. *Cell Biology International Reports* **15**, 75-84 (1991).

140. Bektas,M., Nurten,R., Gurel,Z., Sayers,Z. & Bermek,E. Interactions of Eukaryotic Elongation-Factor-2 with Actin - A Possible Link Between Protein Synthetic Machinery and Cytoskeleton. *Febs Letters* **356**, 89-93 (1994).

141. Bektas,M., Nurten,R., Sayers,Z. & Bermek,E. Interactions of elongation factor 2 with the cytoskeleton and interference with DNase I binding to actin. *European Journal of Biochemistry* **256**, 142-147 (1998).

142. Condeelis,J. Elongation-Factor 1-Alpha, Translation and the Cytoskeleton. *Trends in Biochemical Sciences* **20**, 169-170 (1995).

143. Furukawa,R. *et al.* Elongation factor 1 beta is an actin-binding protein. *Biochimica et Biophysica Acta-General Subjects* **1527**, 130-140 (2001).

144. Ito,T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4569-4574 (2001).

145. Shultz DC. *et al.* Identification of two candidate tumor supressor genes on chromosome 17p13.3. *Cancer Research.* **56(9)**, 1997-2002 (1996).

146. Breuning W. *et al.* Expression of OVCA1, a candidate tumor suppressor, is reduced in tumors and inhibits growth of ovarian cancer cells. *Cancer Research* **59(19)**, 4973-4983 (1999).

147. Classon,M. & Harlow,E. The retinoblastoma tumour suppressor in development and cancer. *Nature Reviews Cancer* **2**, 910-917 (2002).

148. Chen,C.M. & Behringer,R.R. OVCA1: tumor suppressor gene. *Current Opinion in Genetics & Development* **15**, 49-54 (2005).

149. Miller,R.W., Fraumeni,J.F., Jr. & MANNING,M.D. ASSOCIATION OF WILMS'S TUMOR WITH ANIRIDIA, HEMIHYPERTROPHY AND OTHER CONGENITAL MALFORMATIONS. *N. Engl. J. Med.* **270**, 922-927 (1964).

150. Call,K.M. *et al.* Isolation and Characterization of A Zinc Finger Polypeptide Gene at the Human Chromosome-11 Wilms Tumor Locus. *Cell* **60**, 509-520 (1990).

151. Ton,C.C.T. *et al.* Positional Cloning and Characterization of A Paired Box-Containing and Homeobox-Containing Gene from the Aniridia Region. *Cell* **67**, 1059-1074 (1991).

152. Fraumeni,J.F.J.a.G.A.G. **Wilms' tumor and congenital aniridia**. J.A.M.A 206, 825-828. 1968.

153. Hanson,I. & vanHeyningen,V. Pax6 - More Than Meets the Eye. *Trends in Genetics* **11**, 268-272 (1995).

154. Schwartz,F., Neve,R., Eisenman,R., Gessler,M. & Bruns,G. A Wagr Region Gene Between Pax-6 and Fshb Expressed in Fetal Brain. *Human Genetics* **94**, 658-664 (1994).

155. Malpuech,G. *et al.* Male Pseudohermaphroditism, Partial Androgen Receptors Defect, 11P13 Deletion - Indication of Gene Localization. *American Journal of Medical Genetics* **24**, 679-684 (1986).

156. Crolla,J.A. *et al.* A FISH approach to defining the extent and possible clinical significance of deletions at the WAGR locus. *Journal of Medical Genetics* **34**, 207-212 (1997).

157. Crolla,J.A. & van Heyningen,V. Frequent chromosome aberrations revealed by molecular cytogenetic studies in patients with aniridia. *American Journal of Human Genetics* **71**, 1138-1149 (2002).

158. Kimata,Y. & Kohno,K. Elongation factor 2 mutants deficient in diphthamide formation show temperature-sensitive cell growth. *J. Biol. Chem.* **269**, 13497-13501 (1994).

159. Jorgensen,R. *et al.* Exotoxin A-eEF2 complex structure indicates ADP ribosylation by ribosome mimicry. *Nature* **436**, 979-984 (2005).

160. Spahn, CM *et al.* Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J* **23(5)**, 1008-1019 (2004).

161. Lee,H. & Iglewski,W.J. Cellular ADP-ribosyltransferase with the same mechanism of action as diphtheria toxin and Pseudomonas toxin A. *Proc. Natl. Acad. Sci. U. S. A* **81**, 2703-2707 (1984).

162. Sitikov,A.S., Davydova,E.K. & Ovchinnikov,L.P. Endogenous Adp-Ribosylation of Elongation Factor-Ii in Polyribosome Fraction of Rabbit Reticulocytes. *Febs Letters* **176**, 261-263 (1984).

163. Galicka,A., Sredzinska,K. & Gindzienski,A. Cytoplasmic inhibitor of eEF-2 ADP-ribosylation catalyzed by diphtheria toxin or endogenous transferase in rat liver cells. *Biochemical and Biophysical Research Communications* **269**, 553-556 (2000).

164. Bektas,M., Guncer,B., Guven,C., Nurten,R. & Bermek,E. Actin - an inhibitor of eukaryotic elongation factor activities. *Biochemical and Biophysical Research Communications* **317**, 1061-1066 (2004).

165. Izawa,I. *et al.* Identification of Mrj, a DnaJ/Hsp40 family protein, as a keratin 8/18 filament regulatory protein. *J. Biol. Chem.* **275**, 34521-34527 (2000).

166. Winkler,G.S., Kristjuhan,A., Erdjument-Bromage,H., Tempst,P. & Svejstrup,J.Q. Elongator is a histone H3 and H4 acetyltransferase important for

normal histone acetylation levels in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 3517-3522 (2002).

167. Jablonowski,D., Fichtner,L., Stark,M.J.R. & Schaffrath,R. The yeast elongator histone acetylase requires Sit4-dependent dephosphorylation for toxin-target capacity. *Molecular Biology of the Cell* **15**, 1459-1469 (2004).

168. Thompson,J.D., Higgins,D.G. & Gibson,T.J. Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Research* **22**, 4673-4680 (1994).

169. Clamp,M. *et al.* Ensembl 2002: accommodating comparative genomics. *Nucleic Acids Research* **31**, 38-42 (2003).

170. Nicholas et al. The GneDoc multiple alignment viewer. Unpublished . 2005.

171. Bateman,A. *et al.* The Pfam protein families database. *Nucleic Acids Research* **28**, 263-266 (2000).

172. Schultz,J., Milpetz,F., Bork,P. & Ponting,C.P. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 5857-5864 (1998).

173. Letunic,I. *et al.* SMART 4.0: towards genomic data integration. *Nucleic Acids Research* **32**, D142-D144 (2004).